AD-A065 744 PURDUE UNIV LAFAYETTE IND SCHOOL OF ELECTRICAL ENGI--ETC F/G 17/2
ENGINEERING FUNDAMENTALS FOR BASE WIRE TRANSMISSION.(U)
OCT 78 H K THAPAR, B J LEON, T H WEAVER F30602-75-C-0082
UNCLASSIFIED 1842 EEG-EEIC-TR-79-7 NL

1 OF 3
AD
A065744

1842 EEG/EEIC TR 79-7

LEVEL II

TECHNICAL REPORT

ENGINEERING FUNDAMENTALS

FOR

BASE WIRE TRANSMISSION

APPROVED FOR PUBLIC RELEASE - DISTRIBUTION UNLIMITED

BASE COMMUNICATIONS SYSTEMS BRANCH
1842 ELECTRONICS ENGINEERING GROUP (AFCS)
SCOTT AIR FORCE BASE, ILLINOIS 62225
1 MARCH 1979

1842 ELECTRONICS ENGINEERING GROUP

MISSION

The 1842 Electronics Engineering Group (EEG) is organized as an independent group reporting directly to the Commander, Air Force Communications Service (AFCS) with the mission to provide communications-electronics-meteorological (CEM) systems engineering and consultive engineering for AFCS. In this respect, 1842 EEG responsibilities include: Developing engineering and installation standards for use in planning, programming, procuring, engineering, installing and testing CEM systems, facilities and equipment; performance of systems engineering of CEM requirements that must operate as a system or in a system environment; operation of a specialized Digital Network System Facility to analyze and evaluate new digital technology for application to the Defense Communications System (DCS) and other special purpose systems and equipment configuration to check out and validate engineering-installation standards and new installation techniques; providing consultive CEM engineering assistance to HQ AFCS, AFCS Areas, MAJCOMs, DOD and other government agencies.

DISCLAIMER

The use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software; this document may not be cited for purposes of advertisement.

1842 EEG/EEIC TR 79-7

TECHNICAL REPORT
ENGINEERING FUNDAMENTALS
FOR
BASE WIRE TRANSMISSION

BASE COMMUNICATIONS SYSTEMS BRANCH
1842 ELECTRONICS ENGINEERING GROUP (AFCS)
SCOTT AIR FORCE BASE, ILLINOIS 62225
1 MARCH 1979

REVIEW AND APPROVAL

This technical report has been reviewed and is approved for publication and distribution.

GERALD T. HARRIS, Chief
Electronics/Base Sys Engrg Div

ROBERT W. NEILL, Chief
Base Comm Sys Branch

THOMAS H. WEAVER, JR, TAM
Base Teleprocessing Systems

19 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
18 1. REPORT NUMBER 1842 EEG/EEIC-TR-79-7	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 Engineering Fundamentals for Base Wire Transmission		5. TYPE OF REPORT & PERIOD COVERED 9 Final rept., 6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) 10 H. K. Thapar, B. J. Leon, T. H. Weaver, Jr		8. CONTRACT OR GRANT NUMBER(s) 15 F30602-75-C-0082
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Electrical Engineering Purdue University West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Base Comm Systems Branch 1842 Electronics Engineering Group Scott AFB, IL 62225		12. REPORT DATE 11 31 Oct 78 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Rome Air Development Center Griffiss AFB, NY		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release -- Distribution unlimited

12 274 p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

N/A

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Analog transmission
Digital transmission
Analog/Digital conversion
Multiplexing
Concentration

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The base telephone cable plant on each base represents a substantial investment; it was constructed to accommodate the base telephone system and a limited amount of low-speed data. Recently, however, more demanding requirements for the transmission of high-speed data are being imposed on this cable plant in increasing quantities. In order to accommodate these more demanding requirements, we must make more efficient use of the existing cable plant; i.e., utilize it for wider bandwidths than the nominal 4KHz voice channel. → next page (over)

292 000

As a first step, this report reviews the engineering fundamentals involved in the transmission of information over base wire paths. It reviews basic transmission fundamentals, transmission of voice information, transmission of digital data, and multiplexing and concentration techniques.

PREFACE

The AFCS/EPE TR 75-4, Technical Report, "An Evolutionary Plan for Implementation of the Base Level Information Transfer System," (T. Weaver and D. Metters), argued that an evolutionary concept is necessary for the transition from an all analog to an all digital mode of operation on the typical Air Force Base; it also proposed that we make maximum use of existing assets since funding for a one-time all encompassing replacement of those assets would be forthcoming.

The 1842EEG/EEIC TR 78-5, Technical Report, "An Engineering Assessment Toward Economic, Feasibile and Responsive Base-Level Telecommunications Through the 1980's," (T. Weaver), concluded that, through improved maintenance, the use of off-shelf technology (e.g., digital line driver/detector devices and multiplex carrier and concentration devices) and early participation in the digital data network/system design process, the base telephone cable plant can support the majority of new digital requirements through the 1980's.

This report is the first in a proposed series of reports which will provide the technical information necessary to the AFCS base-wire engineer to carry out the goal of making maximum and cost effective use of the base telephone cable plant. It covers the basic engineering fundamentals involved in the transmission of information over base-wire paths.

Following technical reports will cover, base data network engineering, direct digital transmission on cable pairs, multiplex/carrier versus new cable construction analysis, fault-isolation for base data nets and timing distribution for synchronous base systems.

Thomas H. Weaver Jr

T. H. WEAVER, JR., TAM
Base Teleprocessing Systems

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SECTION I

INTRODUCTION

1-1. GENERAL.

A telecommunication network is made up of the following interacting systems: transmission systems, switching systems, and signalling systems. Through the use of these systems, telecommunication networks provide a wide variety of communication services, ranging from the simple call completion between two people to the transmission of such signals as telegraph, data, television, and radio. Transmission systems comprise circuits which carry information between switching centers, between a switching center and a subscriber, or between subscribers; switching systems provide the switching between circuits necessary for "call" completion; signalling systems provide the signals necessary for the control and operation of the switching centers. Thus, due to this interaction between the three systems, it is important that they be properly designed and maintained to assure a successful operation of the entire telecommunication network.

1-2. SCOPE.

This report is intended to provide the basic definitions and concepts encountered in dealing with transmission systems, and more specifically, the transmission systems on the Air Force base. Fundamental concepts about analog and digital transmission are presented.

1-3. ORGANIZATION OF THE REPORT.

Section II gives a detailed description of the fundamental concepts of transmission. These include: basic transmission systems, limitation in transmission, signal measurement, various types of noise, noise measurement, and distortion of signals.

Section III presents the fundamentals of telephone transmission, and includes the following: telephone set, local loop, trunking facility, 2- and 4-wire circuits, switching, transmission engineering standards for the AUTOVON network, distortion in telephone transmission, and introduces the transmission of data on telephone lines.

Section IV presents the fundamentals of digital transmission. These include: analog vs. digital transmission, transmission of data, digital data characteristics, coding, error detection and correction, and synchronous and asynchronous modes of transmission.

Section V presents the various modulation schemes used and their performance.

Section VI deals with the various multiplexing and concentration schemes.

Section VII gives some aspects of network planning and the impact of transmission engineering on the network.

SECTION II

FUNDAMENTALS OF TRANSMISSION

2-1. GENERAL

This section presents the fundamentals of transmission which include the following topics: elements of a transmission system, limitations in a transmission system, types of message transmission, common types of noise, noise and signal measurements, and effective transmission. The terminology introduced in this section forms the basis for understanding the material to be presented in subsequent sections.

2-2. ELEMENTS OF A TRANSMISSION SYSTEM

The physical process whereby a message is transferred from one location, called the source, to another location, the destination, is termed transmission. In general, the source and destination are physically separated. The acoustic pressure produced by speech or music, the data originating from a computer, the holes punched on a computer card, the number on a credit card are all examples of messages. Clearly, all messages are not electrical in nature. For transmission on an electrical system, a transducer is required to convert the message into an electrical signal.

An electrical signal is characterized by its strength (magnitude) and shape ("frequency"). In assessing and assuring good transmission quality, it is these characteristics of the signal that are of paramount concern.

A schematic diagram of a complete transmission system is depicted in Figure 2-1. For the sake of clarity, each functional block is shown as a distinct entity; in an actual system, some of the blocks may be lumped together. In the course of signal transmission, some unwanted effects influencing the signal strength and shape are encountered. Among these effects are: attenuation, which reduces the signal strength; distortion and interference, which effects the signal shape. In Figure 2-1, these effects are shown to occur in the transmission medium; in the actual system, howev-

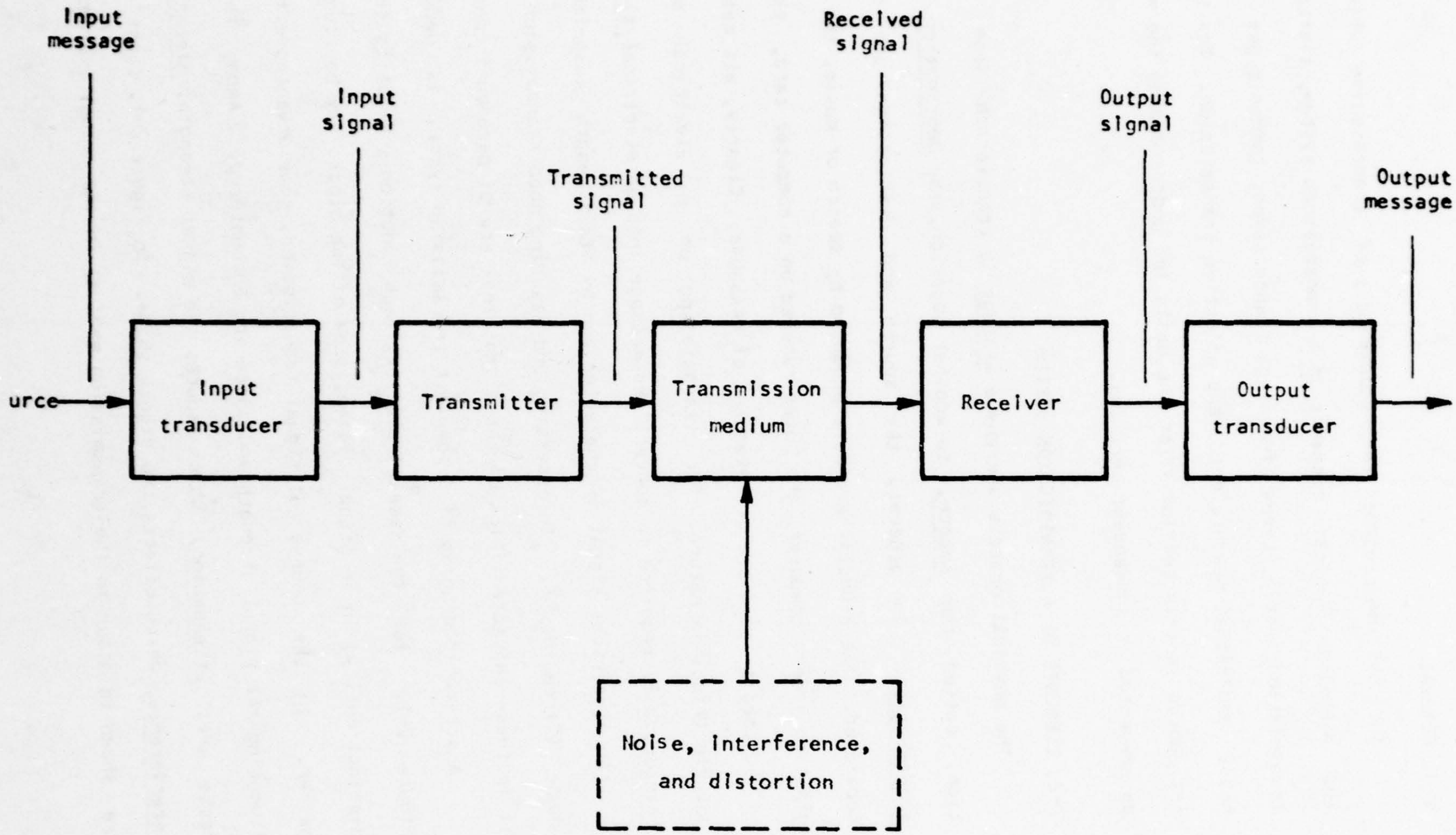


Figure 2-1. Schematic diagram of a Transmission System

er, these may arise in any one of the functional blocks.

2-2.1 INPUT TRANSDUCER. The function of the input transducer is to convert the input message into an electrical signal. Useful transducers include the following: Thermoelectric, which produce an electromotive force (emf) proportional to the temperature and may be used as a temperature sensing device; Piezoelectric, which produce an emf proportional to the applied pressure, and is used in microphones and phonograph pick-ups; Photoelectric, which produce a current proportional to the illumination, and is used in photo-tubes and photocells - for example, a solar cell; Electromagnetic, which produce an emf based on the orientation of the magnetic domains in a magnetic material, and finds applications in tape-recorders and tape-drivers in computer systems. With proper design, circuit elements like resistors, capacitors, and inductors may also be used as transducers.

In telephone sets, carbon granules, which exhibit a change in electrical resistance proportional to the applied pressure, are used to convert voice into an electric signal, and vice versa. The translation of acoustic pressure into an electric signal is an important factor in the quality of telephone (voice) transmission. Some of the common transducers are shown in Figure 2-2.

2-2.2. TRANSMITTER. The transmitter couples the message onto the transmission medium in the form of a transmitted signal. Several signal processing steps, for example, filtering, modulation, coding, amplification, etc., may be performed by the transmitter in order to effectively and efficiently use the capacity of the transmission medium.

2-2.3. TRANSMISSION MEDIUM. The transmission medium, also referred to as the channel, connects the receiver to the transmitter. It may comprise a

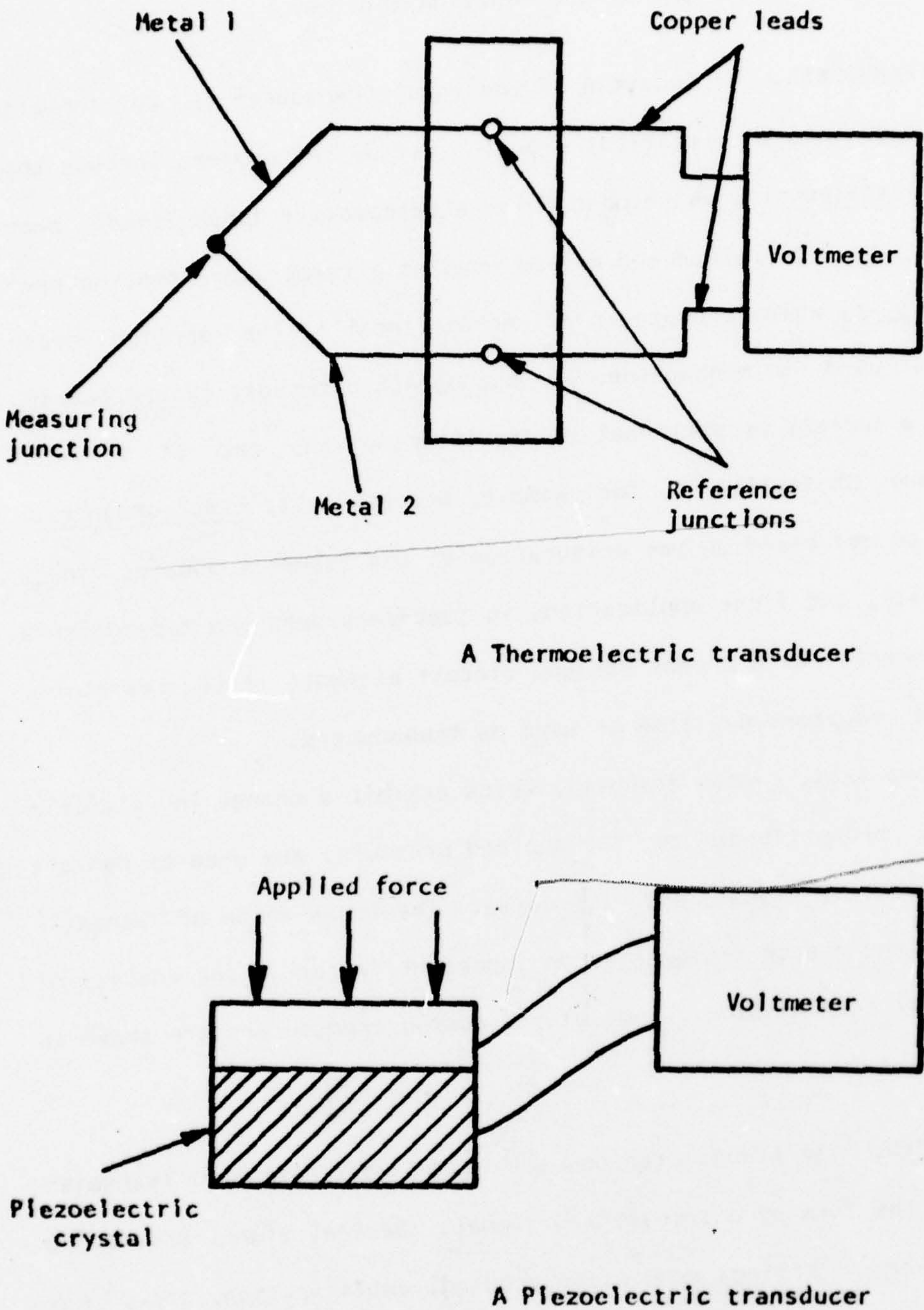


Figure 2-2. Examples of transducers

pair of wires, coaxial cable, waveguide, or a radiating electromagnetic wave such as radio waves and laser beams. A property common to all such media is attenuation, the progressive decrease of signal power with distance.

2-2.4. RECEIVER. The transmitted signal being propagated via the transmission medium is extracted by the receiver for subsequent transfer of the message to its destination. Like the transmitter, the receiver may perform several signal processing operations, which may include amplification, demodulation (inverse of modulation), decoding, etc.

2-2.5. OUTPUT TRANSDUCER. The output transducer converts the output signal from the receiver to the output message. Depending upon the application, the output transducer may or may not be the inverse process of the input transducer. In voice transmission, the output transducer converts the electric signal into speech; thus, the output transducer performs the reverse operation of the input transducer. In data transmission, the input transducer may convert the message in the form of holes on an IBM card into an electric signal, while the output transducer may convert the output signal into an output message to be stored on a magnetic tape. Thus, the function of the output transducer depends upon what form of output message is required, and is not necessarily the reverse of the function of the input transducer.

2-2.6. EXTRANEOUS SIGNALS. Extraneous signals, in the form of noise, distortion, and interference, are inevitable in message transmission. These may arise from natural or man-made causes. Any unpredictable and random electric signal appearing in the various blocks of the transmission system is categorized as noise. Interference is contamination by extraneous signals, which are man-made and predictable. Distortion is the alteration of

the signal due to the inherent imperfections (due to nonlinearities, imbalance, etc.) in the response of the various blocks comprising the system. Unlike interference and noise, distortion disappears when the signal is turned off.

2-3. LIMITATIONS IN TRANSMISSION SYSTEMS

In the design or modification of transmission systems - or any system for that matter - two kinds of constraints are generally encountered: <u>technological problems</u> and <u>fundamental physical limitations</u>.

Technological problems include such considerations as economics, interfacing new systems with existing systems, equipment availability, aesthetics, etc. The knowledge of existing systems and the available options should be at hand when an attempt to solve technological problems is made. For example, in providing AUTOVON services at a base, the available equipment and standards in the DCS AUTOVON Network must be known. To interface the DCS AUTOVON to the commercial system, the operating standards for both systems must be known so that proper interfacing can be accomplished. In general, technological problems are solvable in theory, although the solution may not be feasible in practice.

There is no recourse, however, when fundamental physical limitations are encountered in design. These limitations in transmission system engineering are: Bandwidth and Noise.

2-3.1. BANDWIDTH LIMITATION. All messages, whether voice over a telephone channel, data in a computer communication network, or any other information, are usually transmitted in <u>real time</u>, that is, the output signal keeps pace with the input signal. One aspect of efficient transmission calls for minimizing the message transmission time, that is, sending the most informa-

tion in the least amount of time. In dealing with electrical systems, energy storage is implicit and it takes finite time to change stored energy. Thus, arbitrarily high signal speeds cannot be used because eventually the system will cease to respond in a desirable manner to the signal changes. A measure of signal speed is its bandwidth, the width of the signal frequency spectrum in which the useful energy resides. The rate at which the system can change stored energy in a desirable manner is reflected in the system frequency response, and is measured in terms of the system bandwidth. The bandwidth of the signal being transmitted and the system bandwidth must be compatible: if the system bandwidth is insufficient, decreasing the signal speed, and thereby increasing the message transmission time, may be the only recourse.

The methods for determining bandwidth, and the limitations imposed, differ greatly from system to system. In filters, amplifiers, and attenuators, bandwidth is generally taken to be where a signal will be attenuated by 3 dB below the average level in the passband, or below the level at a reference frequency. For voice telephone signals this reference frequency is 1000 Hz (in the United States) and the bandwidth required for telephone voice is slightly under 4 kHz.

Due to a multifold increase in requirements of communication services around the world, useful bandwidth has become a diminishing resource. Various national and international agencies have been established to allocate bandwidth for different uses. In the U.S. the electromagnetic spectrum is allocated by the Federal Communications Commission. Frequency bands are allocated for radio broadcast, television broadcast, common carrier microwave radio transmission, etc.

2-3.2. NOISE LIMITATION. The accuracy with which the receiver can determine

which signal was actually sent or how well the output device can duplicate the signal from the source is of paramount concern in transmission system engineering. Perfect signal identification or duplication would be possible if there were no noise or extraneous signals arising in the transmission system. But, since the operating temperature of physical systems is not absolute zero, the thermal energy in the system gives rise to a random motion of electrons (apart from other charged particles) which constitutes a random electric current. In conducting materials this produces a random voltage, which is referred to as thermal noise. Also, thermal noise is associated with electromagnetic radiation. Therefore, just as there cannot be electrical transmission without electrons or electromagnetic waves, there cannot be electrical transmission without noise.

The effect of noise on the signal depends upon the strength of the signal at various points in the transmission system. In order to assure satisfactory transmission, precautions must be taken to assure an adequate signal-to-noise ratio (SNR) at every point in the system. It should be noted that if the input signal strength is insufficient, adding more amplifier stages at the receiver end will not improve the signal-to-noise ratio, because the noise will be amplified along with the signal. In fact, the amplifier will add noise to that present at its input.

2-4. TYPES OF MESSAGE TRANSMISSION

2-4.1. SPEECH. The frequency spectrum for speech occupies the 20 Hz to 20 kHz band. The telephone speech signal most commonly encountered has its significant energy concentrated in the 300 to 3300 Hz frequency band. The time-varying waveforms associated with speech signals are not easy to characterize. The audio frequencies comprising the basic speech signal are

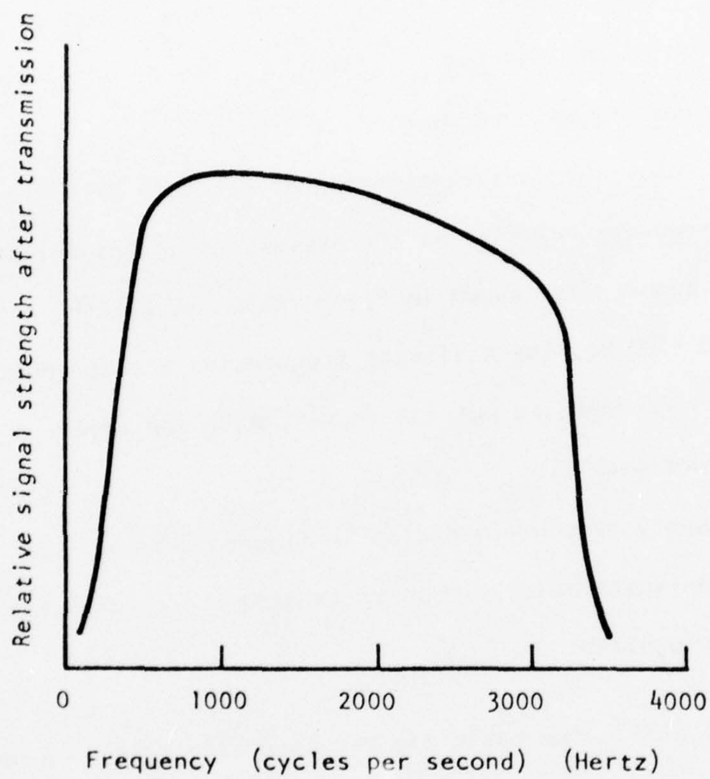


Figure 2-3. Typical frequency response on a voice channel.

varied in amplitude at a syllabic rate. Due to pauses and intonations the speech signal consists of randomly spaced bursts of energy of random duration. Despite this random character, the speech signal must be modeled so that proper design and operation of the transmission system can be undertaken.

The circuit or channel used for transmitting telephone speech signals is referred to as the voice channel. Based on economics, subscriber demands, and quality of transmission, a standard voice channel has a nominal 4 kHz bandwidth. The variation in signal strength with frequency, which is called the frequency response of the system, after transmission over a typical voice channel is shown in Figure 2-3. It will be noted that between about 300 and 3300 Hz, the different frequencies are attenuated by a roughly equal amount. Frequencies outside these limits are severely attenuated, and as such, are not usable.

The frequency response depicted in Figure 2-3 is not sufficient to reproduce human voice exactly, but it is adequate to make the speaker intelligible and recognizable.

2-4.2. DIGITAL DATA. The basic digital data signal is made up of a train of pulses which represent in coded form the data to be transmitted. Various types of data waveforms are discussed in Section IV. The basic waveform consists of frequency components from direct current to some high frequency. Data transmitted on the voice channel is referred to as quasi-analog voiceband data. The speed at which data is transmitted is given in bits per second. Telephone channels, carrying voice-band data, normally transmit at speeds from 45 to 9600 bits per second. Above 1200 bits per second, frequency and phase compensation must be added to the voice channel to provide the minimum values of some line characteristics which are required for data

transmission at the higher bit rates. Such compensation is termed line conditioning.

At relatively low speeds, ranging from 45 to 600 bits per second, Data transmission systems have been developed whereby several of these low-speed data signals are combined to share a single voice channel. These are referred to as voice frequency telegraph (VFTG) systems or voice frequency carrier telegraph (VFCT) systems.

The discussion of data transmission thus far has been in terms of digital data. It should be noted that analog data also exists. For example, signals bearing information regarding temperature, pressure, etc. are generally analog signals. The analog signal may be converted into digital data before transmission, or it may be transmitted the way it is.

2-4.3. WIDEBAND. Signals requiring more bandwidth for transmission than that provided by the voice channel are called wideband signals. Some may be handled by the telephone cable plant while others may require transmission paths consisting of coaxial cable, waveguide or radio facilities. Although used primarily as a voice channel transmission medium, the telephone cable plant is capable of handling up to 1MHZ, depending on the size and length of the wire-pair.

2-4.4. FREQUENCY SPECTRUM The frequency spectra for various types of transmission media are shown in Figure 2-4.

2-5. COMMON TYPES OF NOISE.

Noise, in the broadest sense, is any undesirable signal appearing in the system. A characteristic property of the noise waveform is the random nature of its amplitude variation with time. Accordingly, it is not possible to represent noise on a deterministic basis; representation of its am-

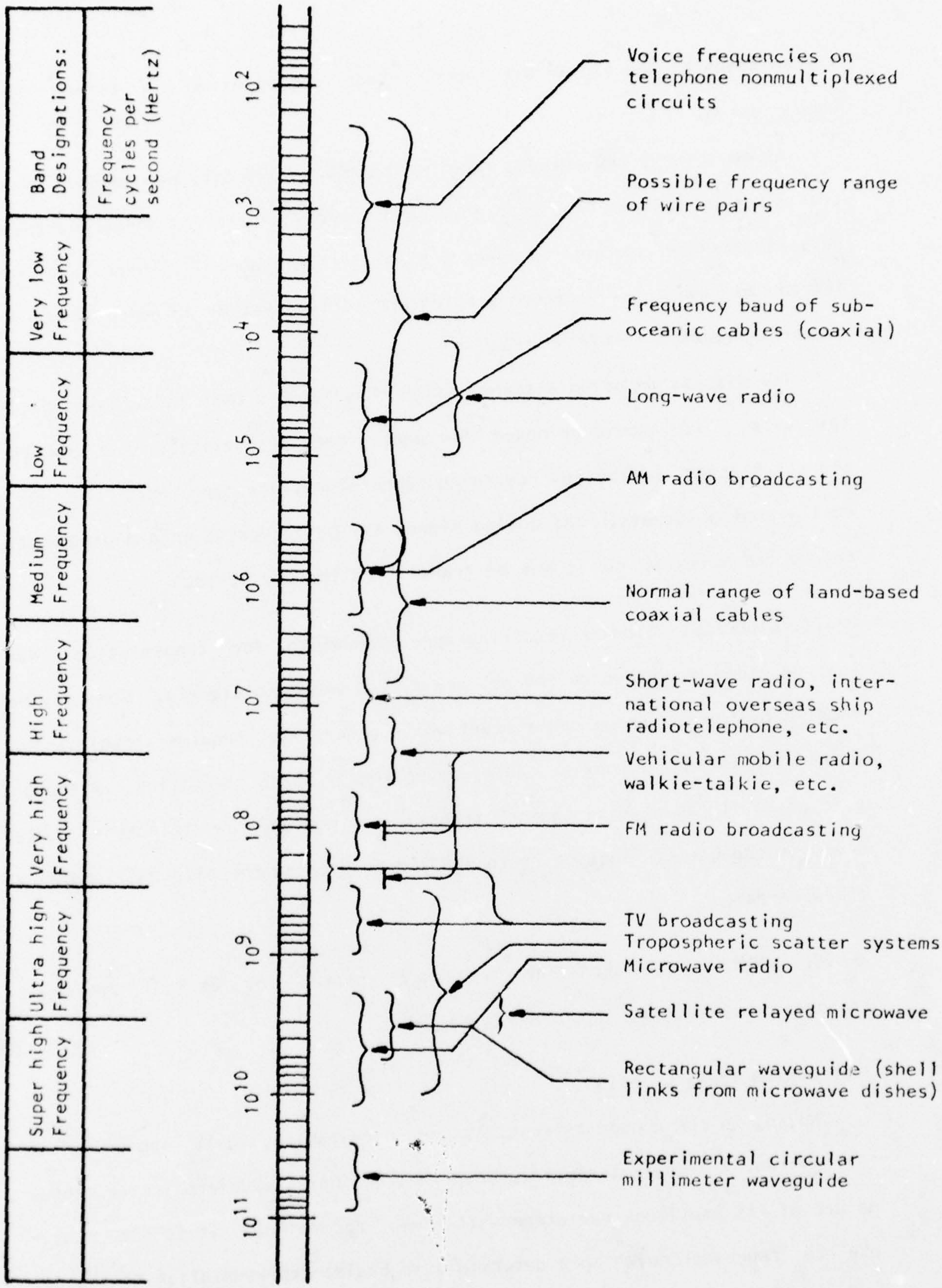


Figure 2.4. Frequency Spectrum in Telecommunications

plitude in terms of probability density functions must suffice. Such signals may be external or internal to the system, and their effect on the system performance depends upon the mode of message transmission and the system design. This effect can be reduced - but never completely eliminated - by taking proper precautions in system design and installation. The need for understanding the types of noise encountered in transmission systems is, therefore, evident.

2-5.1. THERMAL NOISE. Thermal noise is a phenomenon associated with the thermal motion of electrons in conducting media - transmisson lines, resistors, etc. The thermal motion, which is random in nature, gives rise to an a-c component of current; the d-c component is zero (otherwise, an electric field will develop across the conductor due to accumulation of charges at the terminals). The magnitude of thermal noise is proportional to the absolute temperature (in degrees kelvin).

To understand the effect of thermal noise upon a signal being propagated through a system, information regarding its frequency content and amplitude must be known. Thermal noise is characterized as having a uniform distribution of energy over all frequencies (constant power density spectrum). Because of this property, thermal noise is referred to as <u>white noise</u>. White noise, by analogy to white light, has a spectrum which contains all frequency components in equal proportion.

Another important concept pertaining to thermal noise is the <u>Gaussian</u> or <u>normal density function</u>. Thermal noise is due to the random, practically independent motion of a large number of electrons. According to the central limit theorem in probability, when a large number of independent random quantities are summed to form a new random quantity, the distribution function of this new random quantity is Gaussian. A sketch of a typical Gaus-

sian density and distribution functions are shown in Fig. 2-5. Mathematically, the density function is expressed as

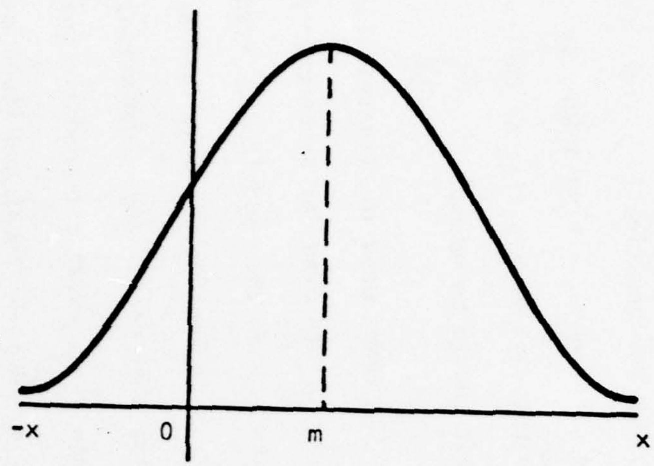
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}(\frac{x-m}{\sigma})^2] \tag{2-1}$$

where x is the random variable, m is the mean, and σ is the standard deviation. The following properties of Gaussian density function should be noted:

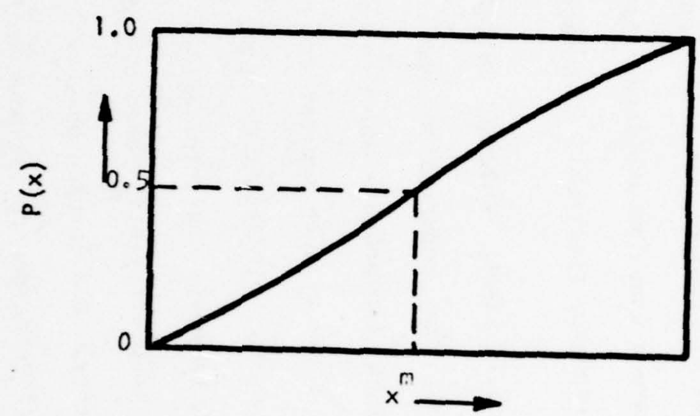
a. From Figure 2-5, it is evident that p(x) describes a continuous random variable, x, that may take on any value from $-\infty$ to $+\infty$, but is most likely to be found near m. This point is called the point of central tendency (point where the group of values tend to cluster).

b. The area under the p(x) curve is unity. From the even symmetry of the p(x) curve, an immediate consequence is that the probability of the random variable taking a value greater than m is equal to the random variable taking on a value less than m.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[- \frac{(x-m)^2}{2\sigma^2} \right]$$



(A) Gaussian density function



(B) Gaussian distribution function

Figure 2-5. Gaussian Density and Distribution Functions

Studies have indicated that thermal noise voltage has a Gaussian distribution. This information helps in ascertaining the probability with which a range of amplitudes occur. The mean square thermal noise voltage is σ^2, and the rms voltage is therefore σ. A table of Gaussian distribution shows that signal magnitudes greater than 3.9σ occur less than 0.01 per cent of the time.

Thus thermal noise is white and Gaussian. However, the terms white and Gaussian are not synonymous: For example, passing Gaussian noise through a linear network will leave it Gaussian, but may drastically alter the frequency content at the output; on the other hand, a single impulse will not have a Gaussian amplitude distribution but will have a flat or white noise frequency spectrum. This important fact should be noted.

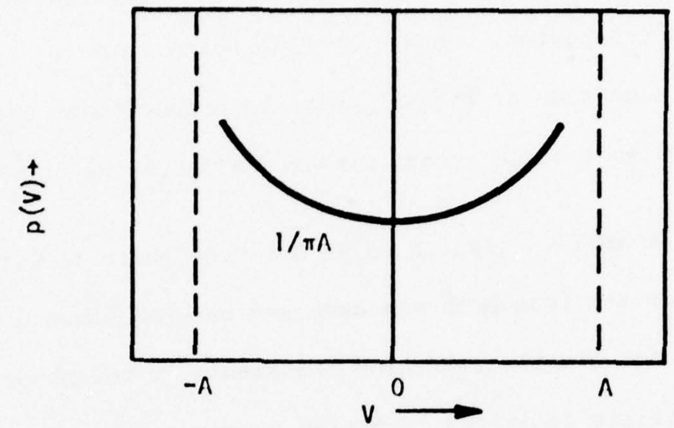
2-5.2. SINGLE FREQUENCY INTERFERENCE. The tones used in transmission systems for supervision and signaling, or pilots used for synchronizing purposes, can produce undesired interference. When the amplitude, frequency, or phase of such sources are random, it is convenient to treat them as noise sources. (It should be noted that in the absence of any randomness, that is, whenever the interference is deterministic, these extraneous signals can be completely eliminated by subtracting a signal of equal amplitude and frequency, and proper phase.) The density function for a single frequency interference source is shown in Figure 2-6. Mathematically, the density is

$$p(V) = \frac{1}{\pi\sqrt{A^2-V^2}}, \quad -A \leq V \leq A \tag{2-2}$$

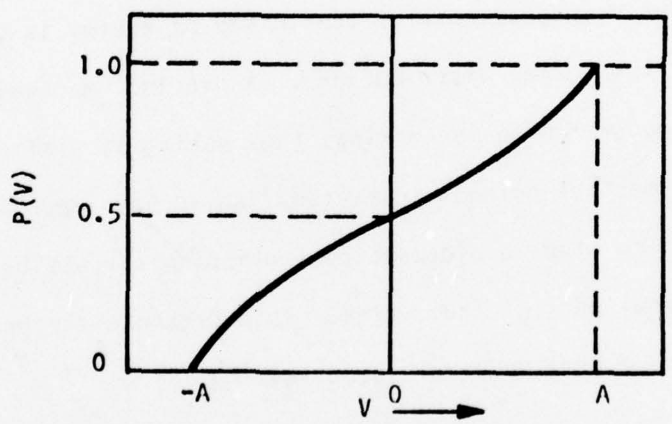
$$= 0 \quad , \quad |V| > A$$

where A is the peak amplitude of the interfering source. The average value

of this noise source is $2A/\pi$, and the rms value is $A/\sqrt{2}$.



(a) Density function



(b) Distribution function

Figure 2-6. Density and Distribution functions for single frequency sinusoidal interference

2-5.3. SHOT NOISE. The discrete nature of electron flow in electrical systems accounts for another type of noise: shot noise. Shot noise is white and Gaussian, and is most common to active devices. It differs from thermal noise due to the following:

a. The magnitude of shot noise is proportional to the amplitude of the signal being propagated, whereas thermal noise is not.

b. The magnitude of thermal noise is proportional to the absolute temperature, but shot noise is not directly affected by it.

2-5.4. IMPULSE NOISE. Impulse noise delivers short spikes of energy to the system. It arises from both man-made and natural causes; some of its common man-made sources are the switching transients in telephone dial central offices and vehicle ignition; lightning is one natural source. The frequency spectrum for impulse noise is approximately flat in the frequency range of interest. The occurence rate of the spikes of energy is often approximated by Poisson or log-normal distributions. A precise mathematical model for analyzing impulse noise is lacking, thus making it difficult to combat its effect on the message being transmitted. An ad hoc approach for reducing its effect is to place a wideband peak-clipping circuit before the bandlimiting (filtering) part of the system. This prevents the pulses from being smeared over more time by simple bandlimiting.

2-6. SIGNAL AND NOISE MEASUREMENTS.

2-6.1. POWER AND VOLTAGE RELATIONS. A transmission system, which comprises transmission lines, amplifiers, filters, etc., can be represented by a cascade of two-port networks. If the input-output relationship of a two-port network can be described by a set of linear differential equations, the network is said to be linear.

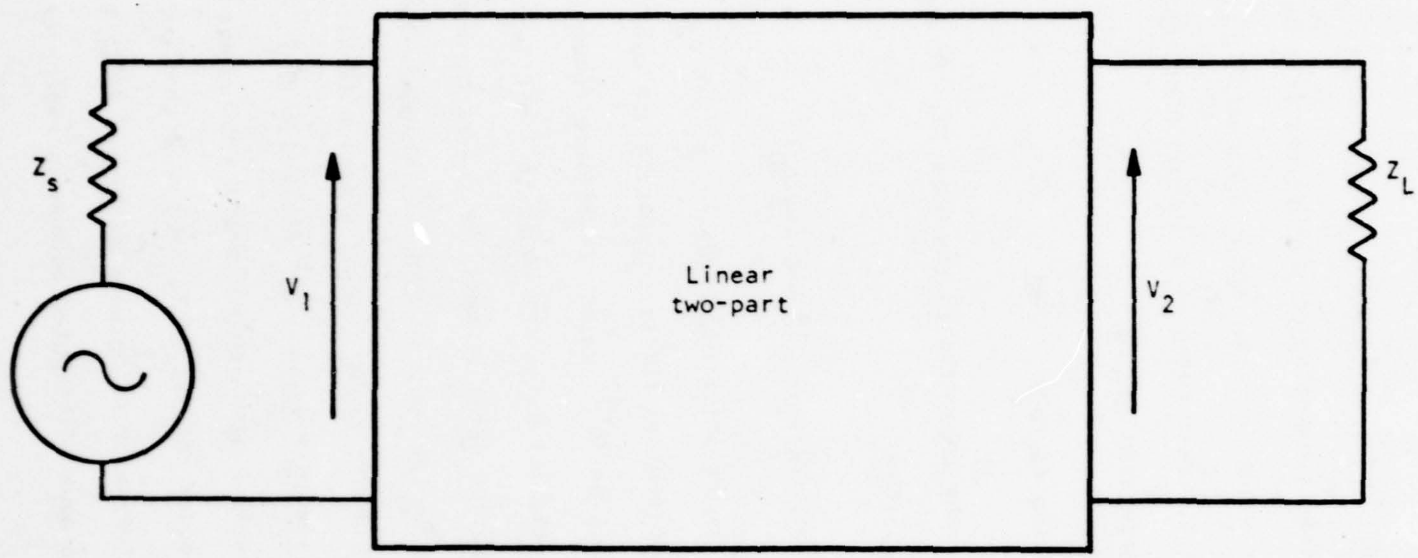


Figure 2-7. A linear terminated two-part.

Consider the linear two-port network shown in Figure 2-7. The input port is driven by a source, V_S, having an internal impedance, Z_S, and the output port has a load, Z_L.

2-6.2. THE DECIBEL. The calculation of the overall gain in multistage amplifiers or systems consisting of sub-systems is simplified by using a logarithmic unit for signal ratios. When two powers, p_1 and p_2, are expressed in the same units, their ratio is a dimensionless quantity. By definition, the decibel (abbreviated dB) is:

$$G = 10 \log (p_1/p_2) \quad \text{dB} \tag{2-3}$$

where log denotes the logarithm to the base 10. If an arbitrary power is represented by p_0, then

$$G = 10 \log (p_1/p_0) - 10 \log (p_2/p_0) \tag{2-4}$$

again represents the relative magnitudes of p_1 and p_2.

In telephone circuits, the signal powers are usually less than 1 watt, and may be as small as 10^{-12} watts. It is more convenient to work with milliwatts (10^{-3} watts). If p_0 is set equal to 1 milliwatt, the terms on the right side of equation (2-4) express the powers p_1 and p_2 in <u>dB relative to 1 milliwatt</u>. This derived unit is abbreviated <u>dBm</u>. If p_0 is set equal to 1 watt, the terms on the right side of equation (2-4) express the powers p_1 and p_2 in <u>dB relative to 1 watt</u>. This derived unit is abbreviated <u>dBW</u>.

The decibel is such a convenient measure that, even though it is defined only for power ratios, it is used as a unit of voltage (or current) ratios. Consider the two-port network of Figure 2-7. When the voltage, V_S, is a sinusoid, the power dissipated across an impedance, $Z_S = R_S + j X_S$, is given by:

$$p_S = \frac{E_S^2}{2R_S(1+X_S^2/R_S^2)} = \frac{1}{2} R_S I_S^2 \qquad (2\text{-}5)$$

where E_S is the amplitude of the sinusoid, and $I_S = E_S/|Z_S|$. Then, if the voltage, V_1, at the input port is $E_1 \cos wt$ and the voltage, V_2, at the output port is $E_2 \cos(wt + \theta)$, then the power ratio in dB is (from equation (2-5)):

$$G = 10 \log \frac{p_1}{p_2}$$

$$= 20 \log \frac{E_1}{E_2} - 10 \log(R_1/R_2) - 10 \log \frac{1+X_1^2/R_1^2}{1+X_2^2/R_2^2} \text{ dB} \qquad (2\text{-}6)$$

$$= 20 \log I_1/I_2 + 10 \log \frac{R_1}{R_2} \text{ dB}$$

In order to avoid any misunderstanding in the usage of dB to express voltage or current ratios, consider the following three cases:

a. When $Z_1 = Z_2$:

$$G = 10 \log \frac{p_1}{p_2} = 20 \log \frac{E_1}{E_2} = 20 \log \frac{I_1}{I_2} \qquad (2\text{-}7)$$

In this case no correction to the reading of a voltmeter calibrated in dB is required.

b. When $Z_1 \neq Z_2$, but $X_1/R_1 = X_2/R_2$: In this case,

$$G = 20 \log \frac{E_1}{E_2} - 10 \log (R_1/R_2) \qquad (2\text{-}8)$$

Voltmeters calibrated in dB will measure only the first term of the right side of equation (2-8). This value should therefore be corrected by subtracting 10 log (R_1/R_2).

c. When $Z_1 \neq Z_2$ and $X_1/R_1 \neq X_2/R_2$:

In this case, the correction factor equal to $(10 \log R_1/R_2 + 10 \log \frac{1+X_1^2/R_1^2}{1+X_2^2/R_2^2})$ should be subtracted from the reading from the voltmeter calibrated in dB.

Cases (a) and (b) are the ones commonly encountered in practice.

For case (a) above, G can be expressed as:

$$G = 20 \log \frac{E_1}{E_0} - 20 \log \frac{E_2}{E_0} \qquad (2\text{-}9)$$

When E_0 is set equal to 1 volt, the terms on the right side of equation (2-9) express the voltages E_1 and E_2 in <u>dB relative to 1 volt</u>. This derived unit is abbreviated <u>dBV</u>. Similarly, when equation (2-9) expresses the voltages E_1 and E_2 in <u>dB relative to 1 mV</u>, the derived unit is abbreviated <u>dBmV</u>.

2-6.3. POWER AND VOLTAGE SUMMATION. One advantage of the decibel unit is that when response is plotted in dB, the overall response curves of a system, consisting of several sub-systems <u>in cascade</u>, can be obtained by adding the individual response curves of each sub-system. In Figure 2-8, the frequency response curves for systems A and B are plotted in dB. The overall response when A and B are cascaded is also shown in Figure 2-8. It is simply the sum of the responses of A and B. Addition of decibels corresponds

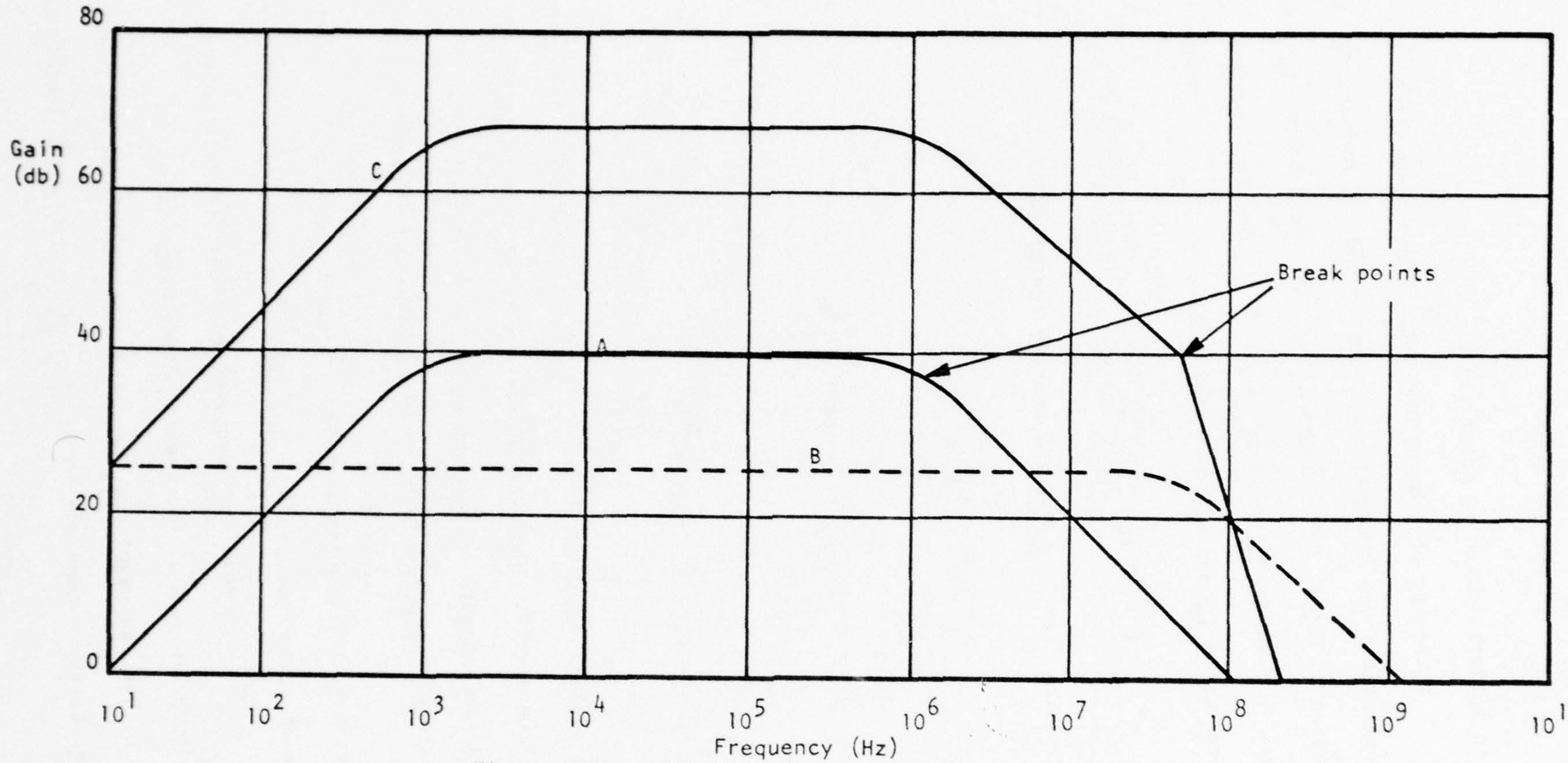


Figure 2-8. Addition of powers in a cascaded system.

(B)

to multiplication of power ratios.

In adding or subtracting two signals expressed in dBm or dBW, the operations are somewhat more complicated. Consider, for example, the system shown in Figure 2-9. Suppose that the input powers p_A and p_B are expressed as P_A and P_B dB, respectively. It is desired to express the output power, the sum of p_A and p_B, as P dBm. Then,

$$P = 10 \log [\log^{-1}(P_A/10) + \log^{-1}(P_B/10)] \qquad (2\text{-}10)$$

Assuming that $P_A \geq P_B$, and rewrite (2-10) as

$$p = p_A(1 + p_B/p_A)$$

then $$P = 10 \log p_A + 10 \log(1 + \frac{p_B}{p_A})$$

$$= P_1 + 10 \log(1 + \frac{p_B}{p_A}) \qquad (2\text{-}11)$$

The maximum value of $10 \log (1 + \frac{p_B}{p_A})$ is about 3dB. This function is plotted in Figure 2-10. When the difference $(P_A - P_B)$ is known, P can be found by adding the ordinate corresponding to the difference $(P_A - P_B)$ to P_A. This graph essentially provides information on how much P exceeds P_A.

Sometimes, the problem is to determine the power due to the sum of two voltage or current waveforms. The following cases are encountered:

a. The two waveforms are uncorrelated: The total power, in this case, is simply the sum of the powers of the components.

b. The two waveforms are correlated: The resulting power can lie anywhere between zero and 3dB more than the power addition result. For exam-

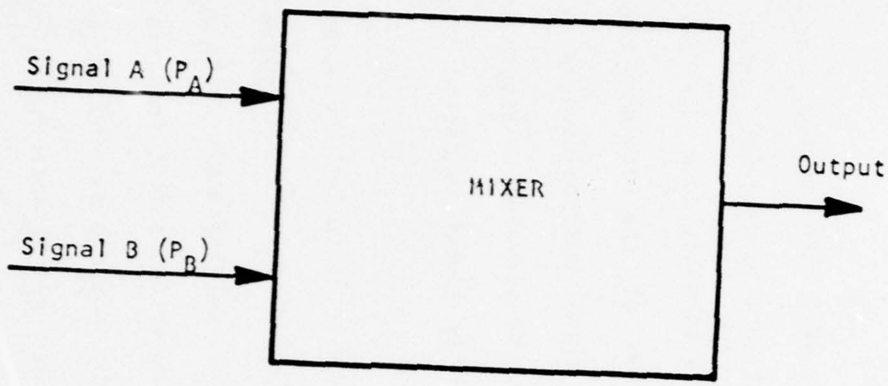


Figure 2-9. Addition of two signals in dB

ple, when two waveforms of equal frequency and amplitude but 180° out-of-phase are added, the resulting power is zero. When the two waveforms are exactly in phase, the resulting voltage is doubled and the power quadrupled. Such an addition is called <u>in-phase addition</u>. A plot for adding two voltage waveforms when their difference in dB is known is shown in Figure 2-11.

2-6.4. VOLUME UNITS (VU). The complex, nonperiodic nature of the speech signal makes it difficult to describe its waveform in terms of rms, peak, or average value - as is commonly done for periodic waveforms. For proper design and operation of the telephone plant, some means of measuring and characterizing this signal must be devised so that gain and loss, overload and distortion can be measured to assure satisfactory system performance. The unit used to measure speech is called <u>volume</u> and is expressed in <u>vu</u> (volume units).

The volume indicator is called the vu meter. The meter scale is logarithmic, with a 10-log scale. The readings bear the same relationship to each other as decibels, but the scale units are still calibrated in vu, and not in dB. The vu meter has a flat frequency response in the audible range and, unlike some of the noise measuring devices, is not frequency weighted in any manner; it measures the approximate rms voltage averaged over a syllabic interval.

2-6.5. TRANSMISSION LEVEL. In a transmission system, comprising a cascade of sub-systems, a string of losses and gains are encountered. To insure satisfactory transmission of the message, it is necessary to account for the signal magnitude at many points relative to its magnitude at other points. A reference, called the <u>transmission level point</u>, is defined to uniquely determine relative signal magnitude at any point in the system. The

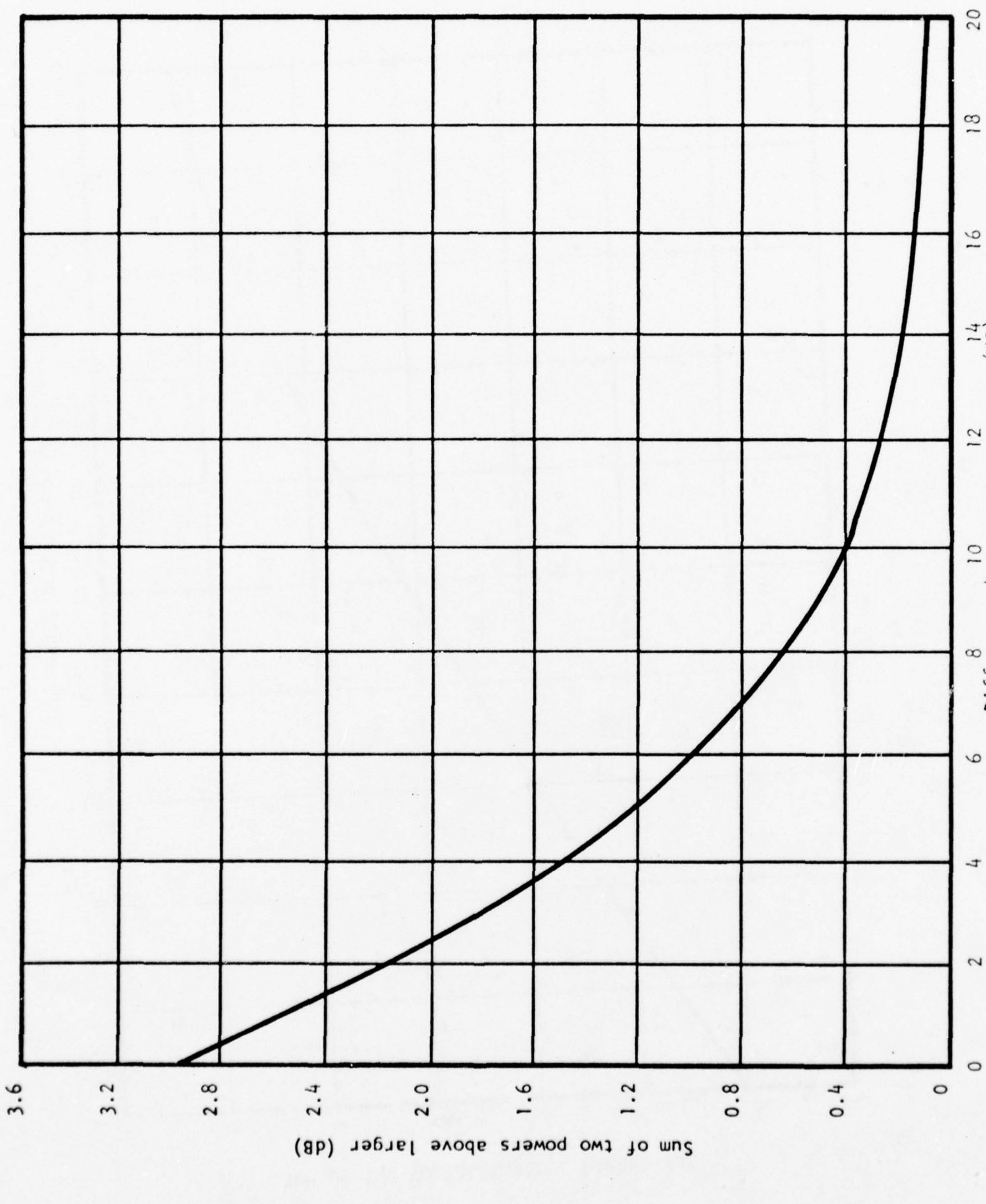


Figure 2-10. Sum of two powers expressed in dB.

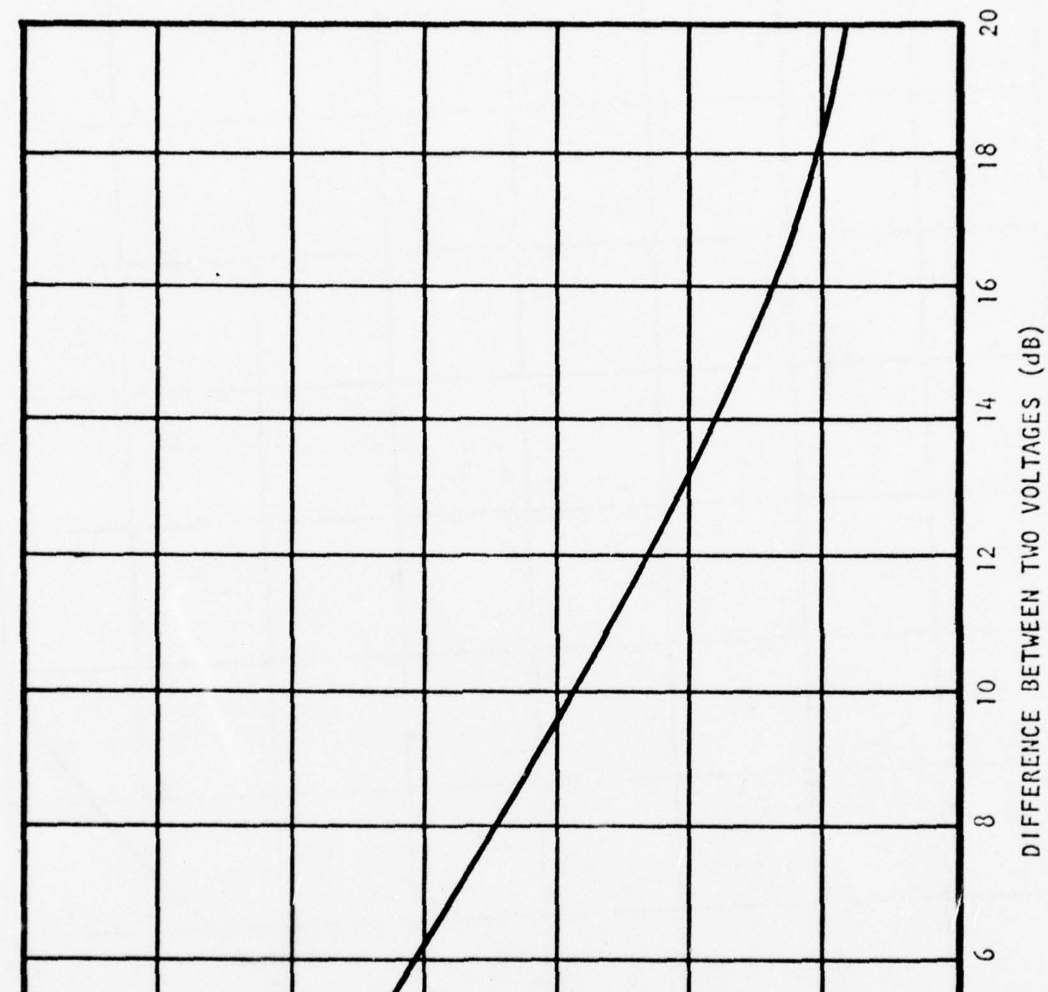


Figure 2-11. Sum of two voltages expressed in dB

transmission level (in dB) of the relative signal magnitude at a particular point is determined by the algebraic summation of the gains, expressed in dB, encountered by the the signal. It should be noted that absolute magnitudes are determined by the applied signal; while the relative magnitudes are determined by the transmission level.

For convenience, this reference point is chosen to be the 0-dB transmission level point. The following terms and abbreviations are commonly encountered in connection with 0-dB transmission level point: zero level, zero level point, zero transmission test level point, 0-dB TL, and 0TLP. Since transmission level at a specific point expresses the ratio of the signal power at that point to the signal power at the reference point, a 0 TLP is, therefore, a point at which the tone level (test tone) is 0 dB (or 0 dBm). Signal magnitude, in dBm, when referred to 0 TLP is expressed in terms of dBm0.

The signal magnitude at any point within the system can be determined when the signal magnitude with respect to the reference level and the level of the point are known. Specifically, if the signal magnitude is S dBm0 at a point whose level is L dB, then the signal magnitude, S_1, at that point is given by:

$$S_1 = S + L \qquad \text{dBm} \tag{2-12}$$

Note

Even though the 0-dB transmission level point is convenient to use on paper, in actual practice, some other transmission level point may be used; this is usually a level that appears at a real interconnect point such as a patch panel.

2-6.6. LOSS, DELAY, AND GAIN. To determine the signal magnitudes at various points within the system, information regarding the transfer of the signal through the networks and sub-systems must be known. To characterize a two-port network, four complex quantities, called the network parameters, are required. In some networks, such as those encountered in telephone transmission, the transfer of the signal from the input to the output can be described by a frequency dependent complex number describing the loss (or gain) and phase shift.

The loss in signal strength and shift in phase of the signal take place predominantly in the transmission medium. The concepts of loss and phase shift will be introduced by showing how a signal is affected as it propagates down a transmission line.

Consider a transmission line with a sinusoidal signal being sent over it, as shown in Figure 2-12. Because of the presence of distributed series resistance of the line, there will be a series voltage drop (=IR) along its entire length. In addition, some of the current will not travel over the entire line and reach the receiver because of the distributed leakage and capacitance between the wires through which the current returns to the transmitter. The magnitude of both the current and the voltage waveforms, therefore, is diminished; and since the power in the line is the product of the current and voltage, the power gradually falls off as the distance from the "generator" (transmitter) increases. Figure 2-12 shows how the amplitude of the voltage, or the current, of the transmitted signal is decreased by the series resistance and the shunt leakage. The transmission line also causes a continuously increasing shift in the phase (phase shift) of the transmitted signal as it travels along the line towards the receiver.

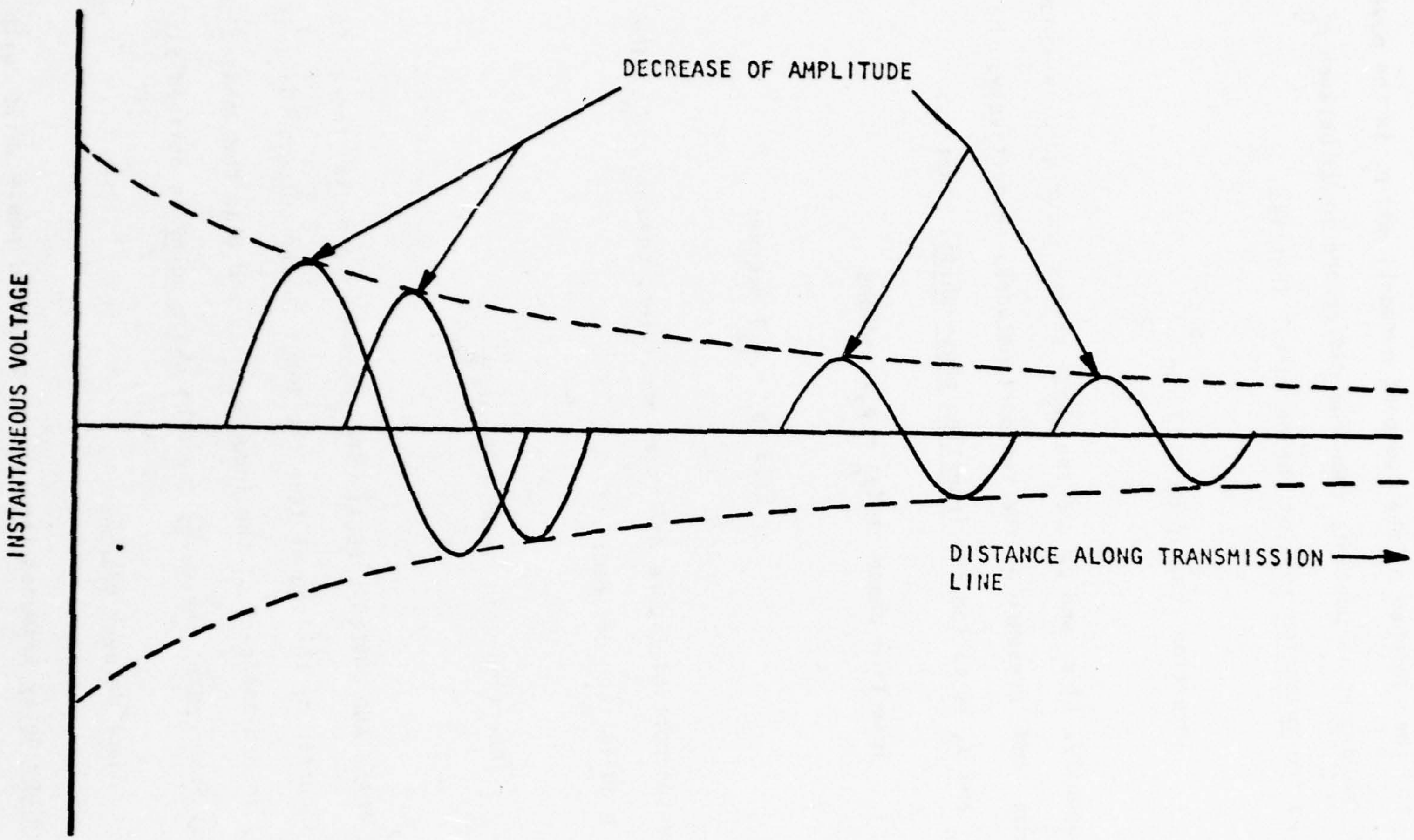


Figure 2-12. Attenuation of Current (or Voltage) Along A Transmission Line

Refer back to Figure 2-7. If p_n is the power delivered to the load, Z_L, in the absence of the two-port network, and p_2 is the power delivered when the 2-port is present, then the difference in dB between p_n and p_2 is called the insertion loss of the network. That is:

$$\text{Insertion loss in dB} = 10 \log \frac{p_n}{p_2} \qquad (2\text{-}13)$$

Analogously, if ϕ_n and ϕ_2 are the phase shifts incurred by a sinusoid in the absence and presence of the two-port network, respectively, the difference in ϕ_n and ϕ_2 is called the insertion phase shift. That is:

$$\text{insertion phase} = (\phi_n - \phi_2) \text{ radians} \qquad (2\text{-}14)$$

$$\simeq 57.3\ (\phi_n - \phi_2) \text{ degrees}$$

If the two-port of Figure 2-7 is an amplifier, then $p_2 > p_n$. The insertion gain in dB is then defined:

$$\text{Insertion gain in dB} = 10 \log \frac{p_2}{p_n} \qquad (2\text{-}15)$$

2-6.7. PHASE AND ENVELOPE DELAY. In transmission media, there is a phase shift incurred by the signal from the input to the output. If ω is the frequency, in radians/sec, of the input signal, and θ is the phase shift, in radians, incurred by the wave, then the phase delay is defined as:

$$\text{phase delay} = \theta/\omega \text{ sec.} \qquad (2\text{-}16)$$

The envelope delay measures the rate of change of phase shift with respect

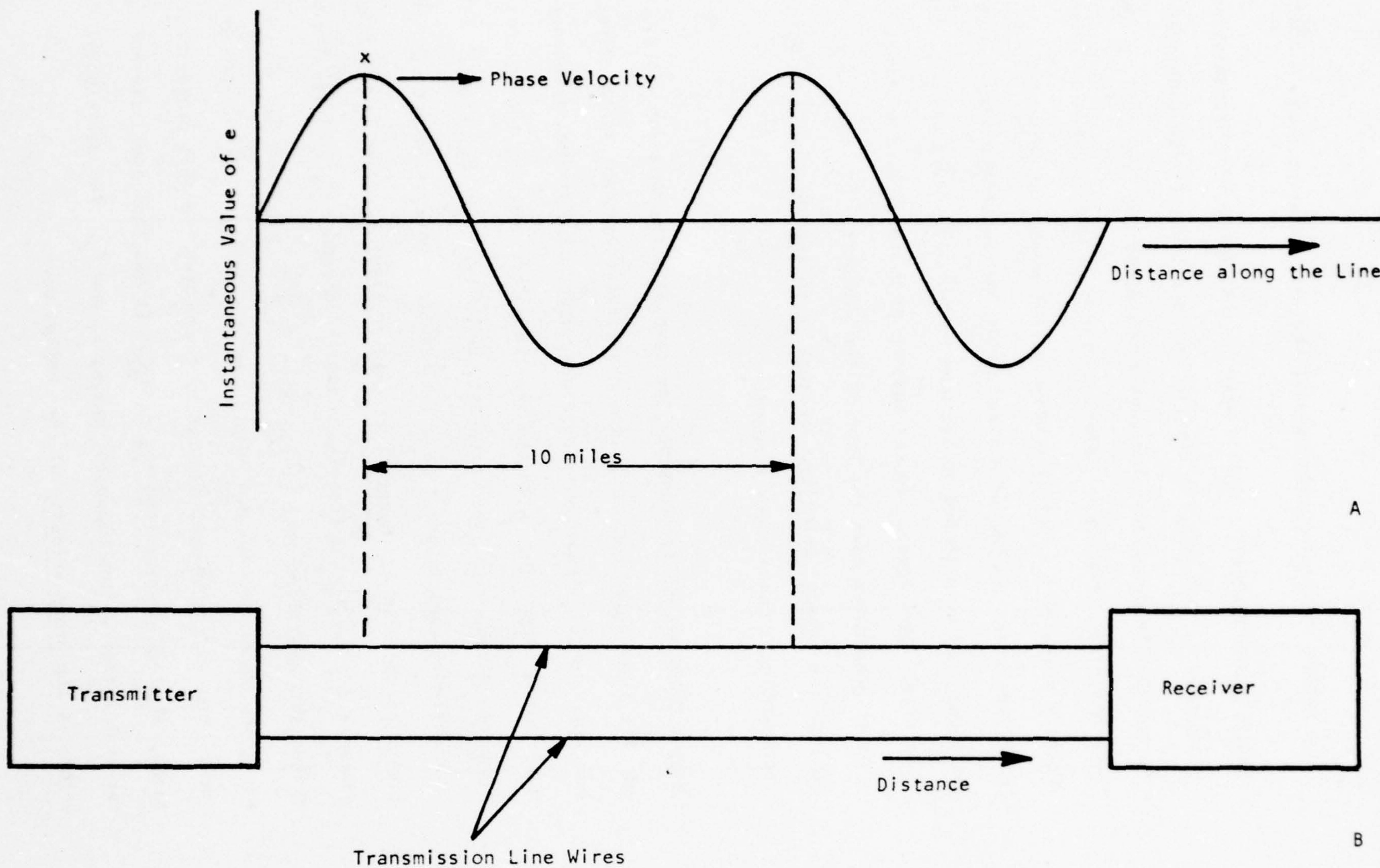


Figure 2-13. Motion of a sine wave illustrating the concept of phase velocity.

to a change in the input frequency; that is, envelope delay = $\frac{d\theta(\omega)}{d\omega}$.

2-6.8. PHASE VELOCITY. When a wave is introduced in a transmission system, especially a transmission line, a very small but finite time goes by before it reaches the receiver. The wave moves along the line with a very high, but definite velocity. Refer to the sine-wave signal in Figure 2-13(A). Point x moves with a definite phase of the wave - in this case, it is the positive peak. x is not a fixed point, but a hypothetical one that should be thought of as attached to the wave itself. It moves down the line as the electrical wave moves. In the absence of distortion, the velocity of point x as it propagates down the line is the velocity of the wave. This is called the phase velocity, because it is the velocity with which a point representing a phase moves forward.

2-6.9. GROUP VELOCITY. Consider the case of a transmission line. Because of its distributed constants, the phase-shift constant is not directly proportional to the frequency. Therefore, waves of different frequency do not move along the line with the same velocity. Rather, the wave of highest frequency moves with the greatest velocity. In B, Figure 2-14, the transmitted waves f_1 and f_2 and the resultant wave, F, start with the relationships shown in A. However, at a later time, t_2, wave f_2 has traveled a distance d_{p2} but f_1 has traveled a shorter distance d_{p1}. Thus, the phase of f_1 has shifted behind that of f_2. When f_1 and f_2 are added, the resultant wave, F, has its maximum amplitude at a point different from that on an ideal line. Its envelope, denoted by the broken line has progressed a distance, d_g, now greater than d_{p1} or d_{p2}. It has thus moved farther and faster than either of the component waves, f_1 and f_2. The velocity of the envelope is the group velocity of the complex wave.

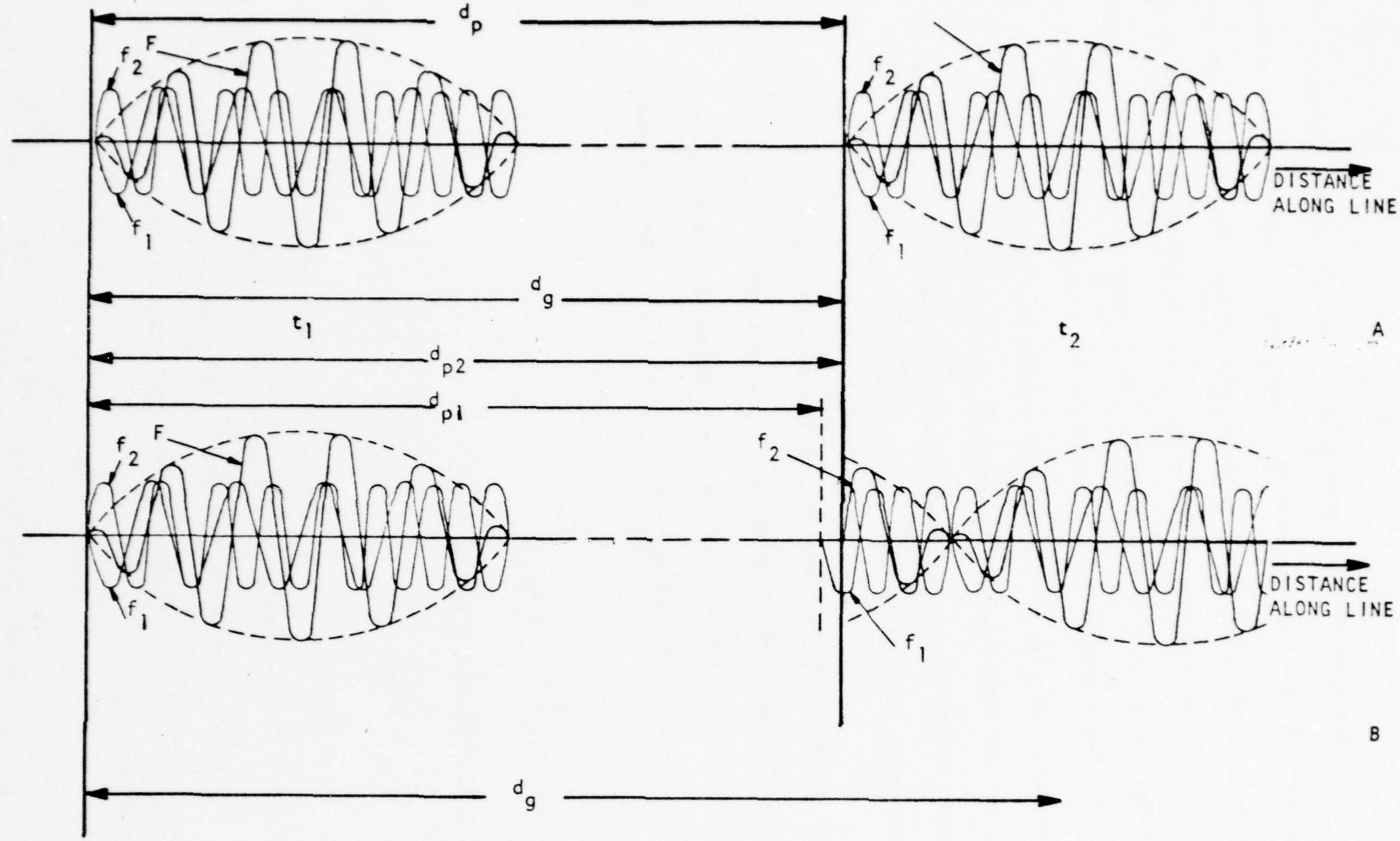


Figure 2-14. Motion of two sine waves illustrating the concept of group velocity.

The group velocity is always greater than the phase velocity of any component of the complex wave on a transmission line (for wave guides, the reverse is true). Although the group velocity may be thought of as the velocity of the envelope of the wave, it is not the velocity with which the intelligence is conveyed by a complex signal, such as an amplitude modulated carrier. The intelligence is conveyed by the sidebands, at a velocity which is the phase velocity of the sideband components. Neither the energy of the wave, nor the intelligence it represents, can travel faster than the fastest component of the wave. For simplicity, the attenuation of the wave has been ignored here; attenuation does occur, however.

2-6.10. AVAILABLE GAIN. Consider, once again, the two-port network of Figure 2-7. Maximum power from the source, with internal impedance Z_S, will be transferred to the two-port if the input impedance of the two-port is Z_S^*, where Z_S^* denotes the conjugate of Z_S; that is, for maximum power transfer, if $Z_S = R_S + j X_S$

then $$Z_{in} = R_S - j X_S \qquad (2\text{-}17)$$

The <u>maximum available power</u> is

$$p_{a1} = E^2/4R_S \qquad (2\text{-}18)$$

where E is the open-circuit generator rms voltage. The power delivered to the load, Z_L, will also be maximized if the output impedance - i.e., the impedance looking in from the output port - is conjugate to Z_L. If p_{a2} is the power delivered to the load under such conditions, then the <u>available gain</u> is defined as:

$$g_a = 10 \log (p_{a2}/p_{a1}) \tag{2-19}$$

2-6.11. POWER GAIN. In practice, the conjugacy requirements are not always met. Then, if p_1 is the actual power delivered to the input-port and p_L is the actual power delivered to the load, then the power gain of the two-port is defined as:

$$g_p = 10 \log(p_L/p_1) \tag{2-20}$$

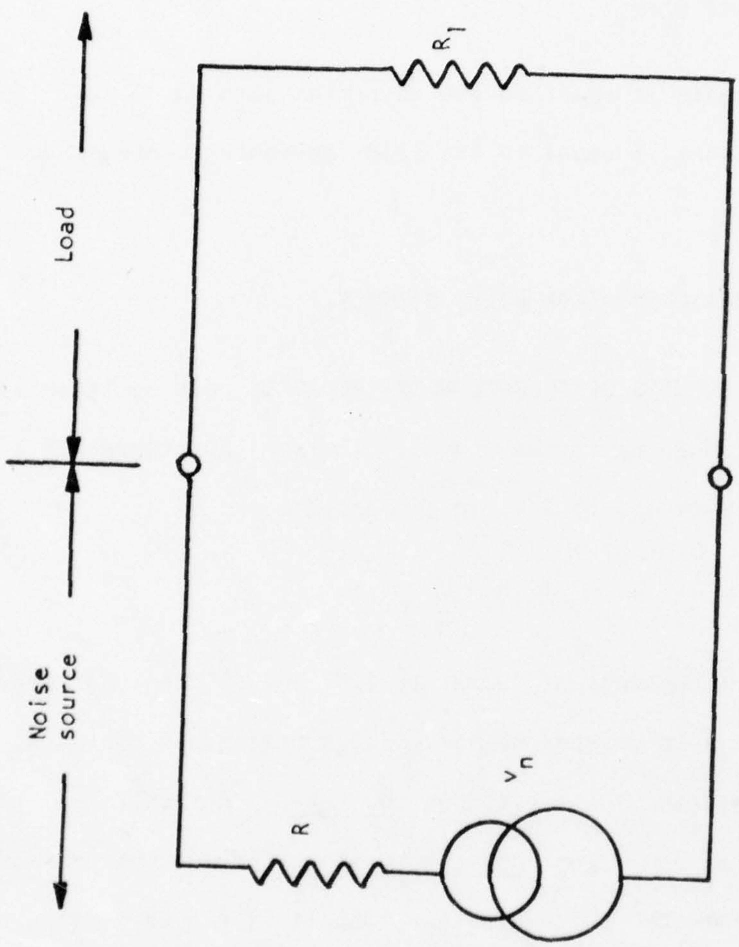
Clearly, the power gain is equal to the insertion gain when the input impedance of the network is equal to the load impedance connected at the output of the network.

2-7. MODELING AND MEASUREMENT OF NOISE SOURCES.

2-7.1. EQUIVALENT CIRCUITS OF THERMAL NOISE-SOURCES. All resistances in a circuit are accompanied by thermal noise. The available power of a thermal noise source at a temperature, T^oK, in a bandwidth, B, is

$$p_a = kTB \tag{2-21}$$

where k is the Boltzman constant, equal to 1.37×10^{-23} ergs per degree, and the temperature T is in degrees kelvin. For modeling and analysis, a noisy resistor can be replaced by a fictitious noiseless resistor of the same value, along with an equivalent noise generator. From equations (2-17) and (2-20), it is evident that the equivalent circuit of Figure 2-15 will suffice to model a noisy resistor, where the mean square generator voltage is proportional to the resistance, to the absolute temperature, and to the frequency bandwidth in which the noise voltage is measured.



$v_{n_{rms}} = \sqrt{4kTRB_w}$

Figure 2-15. Modeling of a noisy resistor.

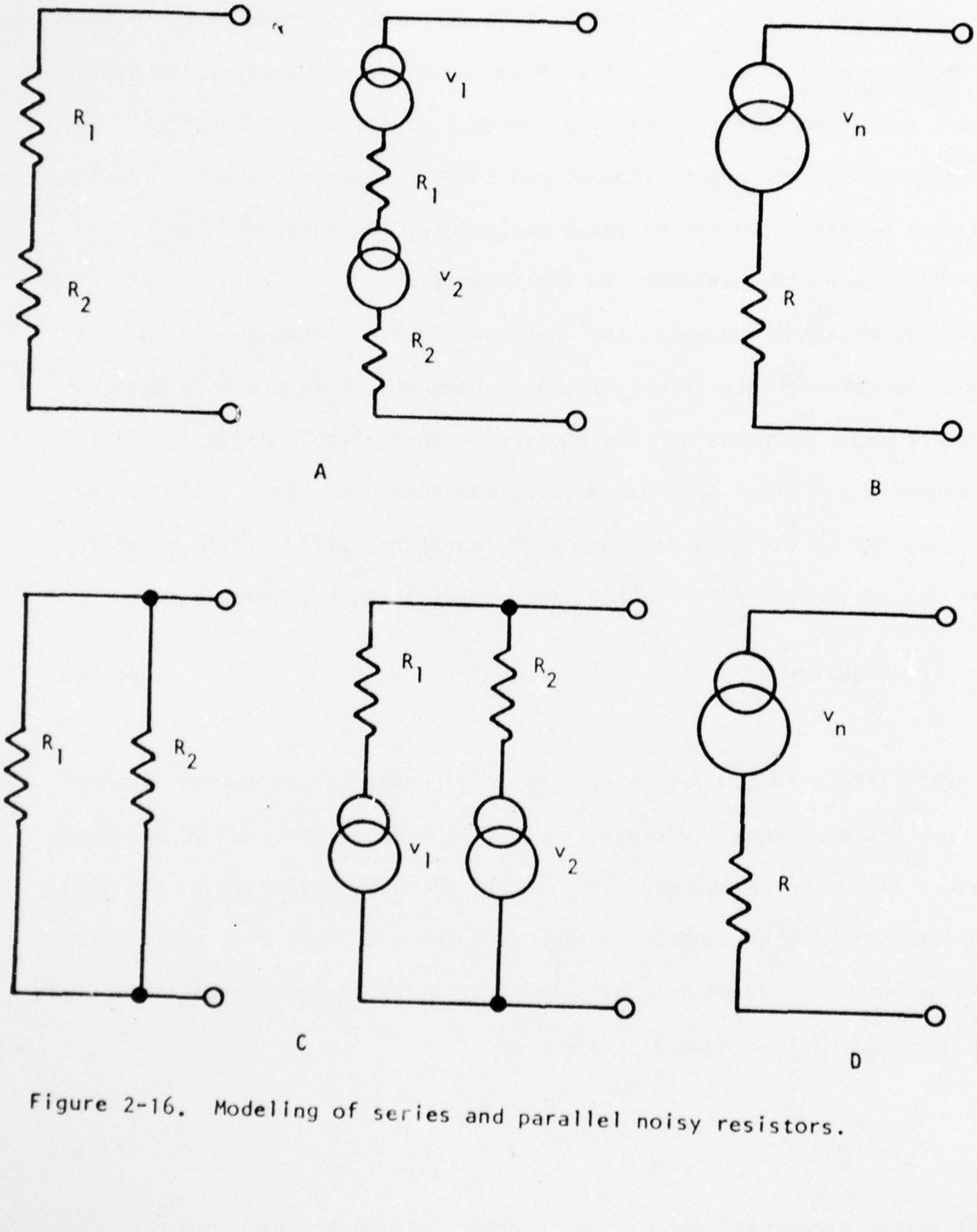


Figure 2-16. Modeling of series and parallel noisy resistors.

When two noisy resistors, R_1 and R_2, are connected in series as shown in A, Figure 2-16, the equivalent circuit shown in B is obtained. When two noisy resistors, R_1 and R_2, are connected in parallel as shown in C, Figure 2-16, the equivalent circuit shown D is obtained. In general, it can be shown that for a two-terminal resistive network having all resistors at the same temperature, the mean square open circuit noise voltage is directly proportional to the value of the input resistance, regardless of the inter-connections between the resistors in the network.

Reactive elements - namely, the inductor and the capacitor - do not contribute to noise in electrical systems. However, their presence does affect the frequency response and the equivalent open circuit voltage of the noise source. For a two terminal device, consisting of passive linear elements, R, L, and C, having a driving point impedance, $Z(f) = R(f) + jX(f)$, the mean square circuit noise voltage is a small frequency band, δf, is

$$e_r^2 = 4kTR(f)\delta f \tag{2-22}$$

2-7.2. NOISE TEMPERATURE. Temperature is the fundamental parameter of thermal noise; the mean square voltage is directly proportional to the absolute temperature of the noise source. The concept of <u>noise temperature</u> is used in characterizing other noise sources - nonthermal, etc. - as well. If a given noise source produces an available power N_a, then the equivalent noise temperature, T_S, of this source is given by:

$$T_S = \frac{N_a}{k\delta f} \tag{2-23}$$

where δf is the frequency interval of interest. It should be noted that the noise temperature for a thermal noise source is its physical temperature;

for any other noise source it is simply a measure of the available noise power, and is not related to its physical temperature. From equation (2-22), it should be evident that noise temperature of a source can be frequency dependent. Also, it should be noted that the concept, and use, of noise temperature is not restricted to noise sources alone; for example, the noise power measured at the output of an amplifier can be characterized in terms of an equivalent noise temperature.

Another concept used in modeling noise sources is excess noise temperature. This concept is especially convenient when handling systems operating at room temperature. Excess noise temperature measures the difference between the noise temperature of the source under consideration and the noise temperature of a thermal noise source at standard or room temperature, which is conventionally taken as 290K.

$$T_e = T_S - T_R \qquad (2\text{-}24)$$

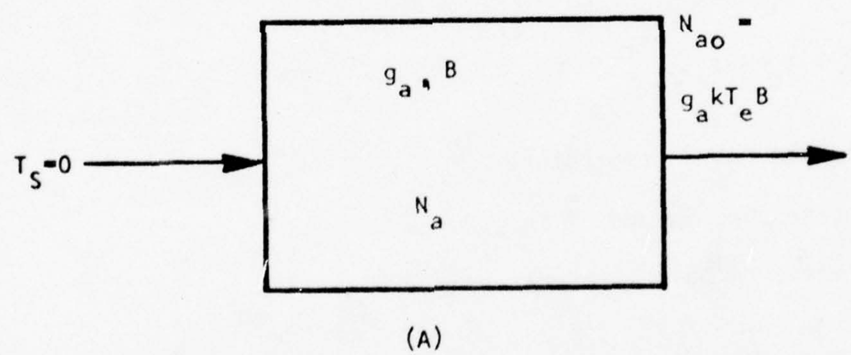
where T_e = excess noise temperature

T_S = noise temperature of the source

T_R = reference noise temperature = 290K = 17°C.

2-7.3. EFFECTIVE INPUT NOISE TEMPERATURE. The noise power at the output of a two-port network consists of two components: the power due to the external noise source (noise along with the input signal) and the power due to the internal noise sources. In a transmission system, consisting of a cascade of two-port networks, it would be cumbersome to treat these sources of noise separately or to repeatedly add the noise powers to compute their cummulative effect. The situation is simplified when the noisy two-port is replaced by an equivalent noise free two-port along with a fictitious noise source at the input such that the noise power characteristics at the output

terminals are retained. The noise power of the fictitious noise source is specified in terms of effective input noise temperature, which is that input source noise temperature which, when connected to the noise free equivalent of the two-port, produces an output noise power equal to that of the actual network connected to a noise free source. The concept of effective input noise temperature is illustrated in Figure 2-17. In A, Figure 2-17, a noisy amplifier with noise free source ($T_S = 0$), and output noise power N_{a0}, is shown. Figure 2-17(B) shows the noisy amplifier replaced by a noise-free amplifier, with the source temperature now equal to T_e. Thus, the effective input noise temperature is a measure of the noise in a two-port when referred to the input terminals. If the device is noiseless, $T_e = 0$.



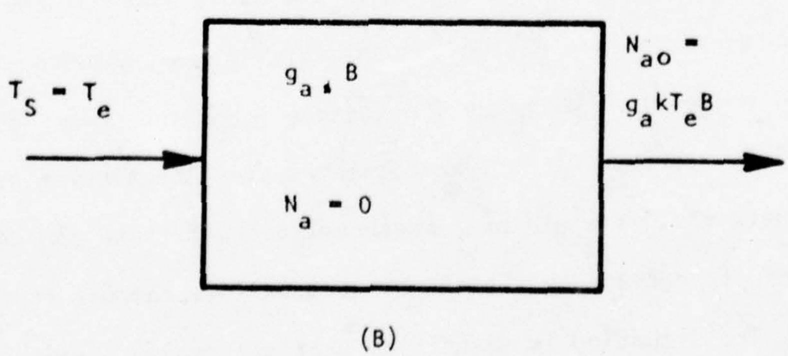


Figure 2-17. Illustration of effective noise temperature concept: A. Noisy amplifier B. Noisefree equivalent.

2-7.4. NOISE FIGURE. Instead of using the effective input noise temperature description, the concept of noise figure is often used to describe a two-port network. Thermal noise, because it is so weak, is of concern when a circuit contains considerable amplification. An absolutely quiet amplifier would amplify equally both the input signal and the thermal noise at the input, and would therefore not change the signal-to-noise ratio. However, amplifiers always contribute some noise of their own. The quietness of a two-port network is measured in terms of its noise figure. The noise figure of a two-port is the number of dB which the noise delivered to an output circuit is greater than the amplified thermal noise from the circuit connected to its input. The term noise factor is applied to the power ratio corresponding to the noise figure. One cause of noise in an amplifier is the thermal noise in its input resistance. Figure 2-18 shows an amplifier having an input resistance of R_1 connected between two circuits of resistance R. If the impedance of the amplifier input is designed to match that of the circuit, that is, if $R_1 = R_2$, the noise in R_1 which is amplified by the full gain of the amplifier increases the noise delivered to the output circuit. The amount of increase will be 3dB. However, if R_1 is made very large, the noise is practically eliminated. Thus, by deliberately mismatching the amplifier impedance to the input line, its noise figure is improved. Noise figure, n_F, is related to effective input noise temperature as follows:

$$n_F = 1 + \frac{T_e}{T_0} \tag{2-25}$$

where T_e is the effective input noise temperature and T_0 is the standard temperature.

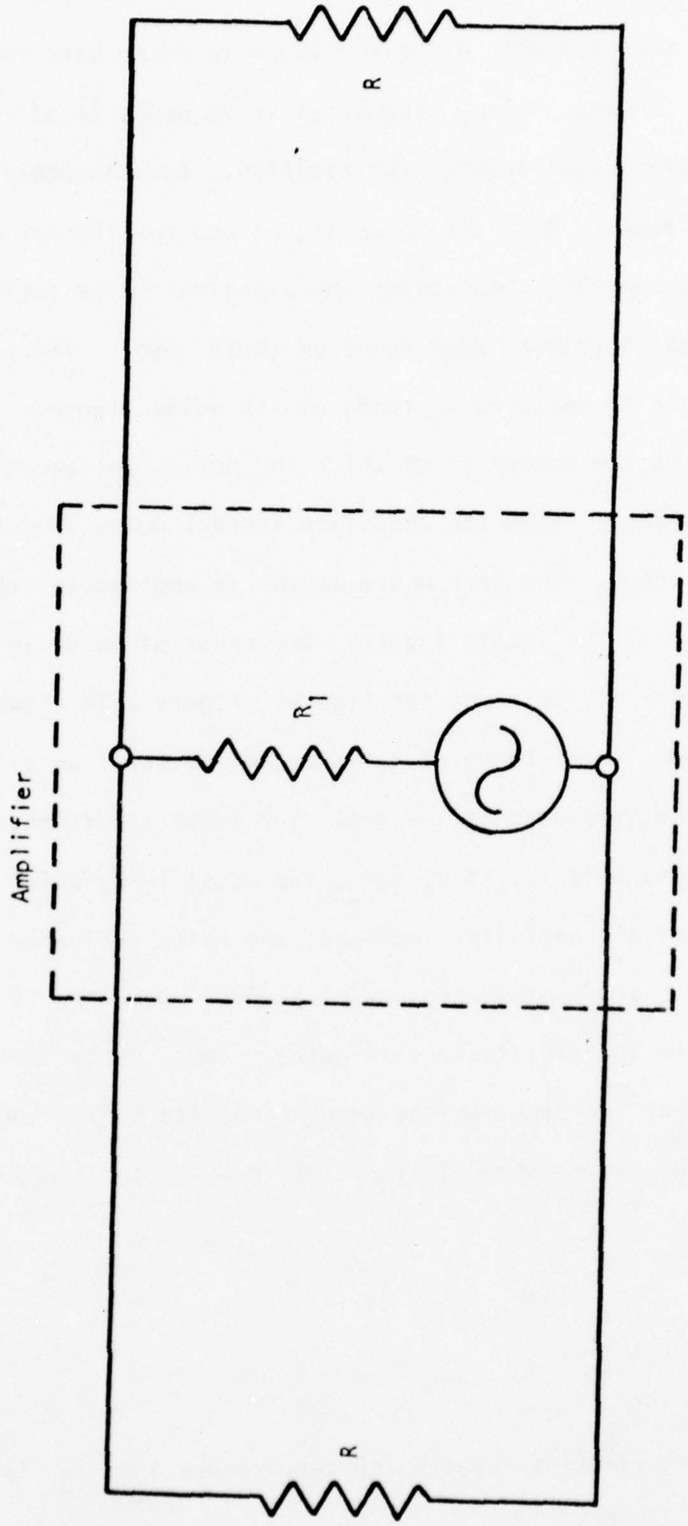


Figure 2-18. Mismatching to Improve Noise Figure

2-7.5. MEASUREMENT OF EFFECTIVE INPUT NOISE TEMPERATURE. The effective input noise temperature of a two-port network can be measured by using the basic arrangement of Figure 2-19. Either one of the following two methods may be used:

a. Calibrated noise source method: In this method, two output power measurements are needed when the input is a calibrated noise source. The first measurement is taken with the noise source turned OFF. The noise temperature, T_1, of the source due to its terminating resistance is known. In the second measurement, the noise source is ON. The noise temperature, T_2, of this source is known. If p_1 and p_2 are the output noise powers in the first and second measurements, respectively, then the effective input temperature of the two-port can be calculated from:

$$T_e = \frac{T_2 - \frac{p_2}{p_1} T_1}{\frac{p_2}{p_1} - 1} \tag{2-26}$$

b. Calibrated input source method: This method also requires two output power measurements to calculate the effective input noise temperature. The input consists of a calibrated signal source whose internal impedance is assumed to be at room temperature. If P_a is the available power of the source, and p_1 and p_2 are the output noise powers with the source OFF and ON, respectively, then the effective input noise temperature of the two-port is given by:

$$T_e = \frac{p_a}{kB(\frac{p_2}{p_1} - 1)} - T_0 \tag{2-27}$$

where T_0 is the absolute room temperature, k is the Boltzman constant, and B

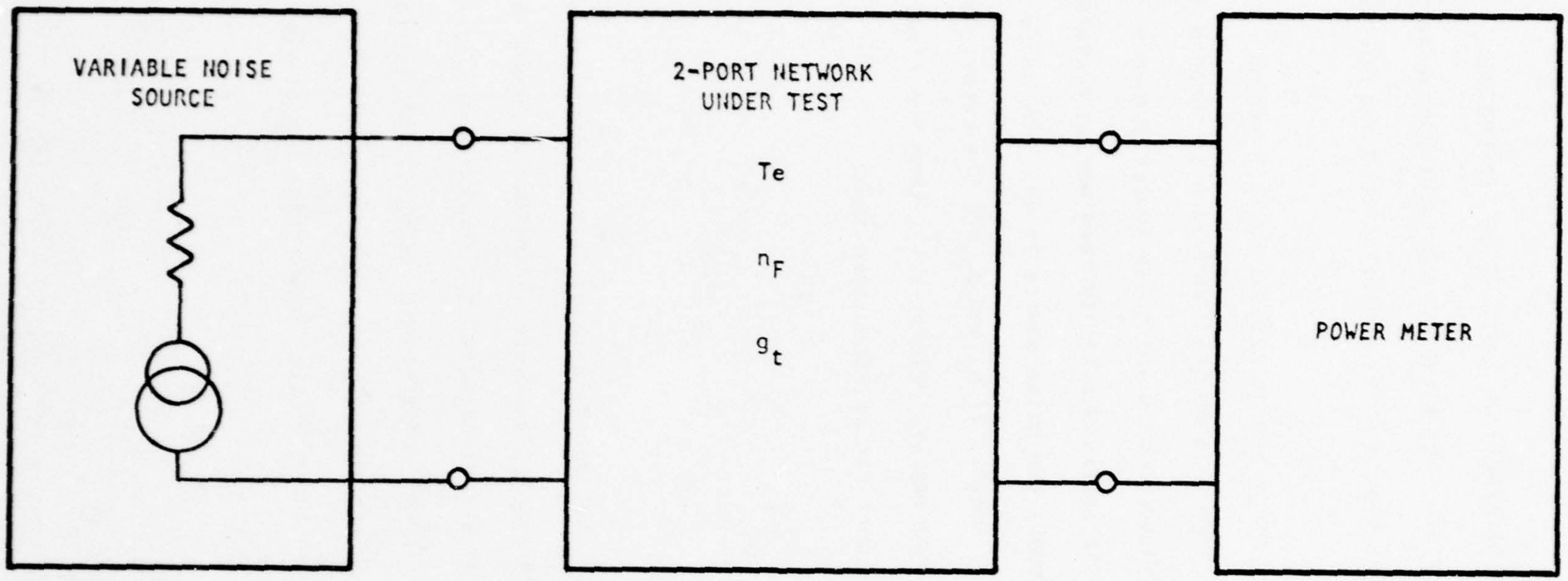


Figure 2-19. Arrangement for measuring effective noise temperature

is the noise bandwidth.

The measurement of the noise bandwidth in equation (2-26) is not easy. Thus, method (a) has an advantage in that the determination of noise bandwidth is not required. It should be noted that Noise Figure can also be determined by using any one of the methods in paragraph (2-7.2) and equation (2-24).

2-7.6. NOISE MEASUREMENT ON TELEPHONE CHANNELS. In the design and operation of voice transmission systems, a knowledge of the absolute magnitude of the average noise power is not enough. The interfering effect of noise on an average telephone subscriber must be known and remedied. This interfering effect can be twofold: its presence can cause annoyance to the subscriber and/or it can render the transmitted message unintelligible. Both these effects are functions of the frequency and magnitude; that is, one frequency may be more disturbing than another even when both have the same magnitude. With a knowledge of the relative disturbing effect of various frequencies, necessary steps to reduce their effect can be taken. By attenuating the various frequencies differently during transmission/receiving, the subsequent interfering effects can be reduced.

a. Noise meters used in voice transmission contain a weighting network that adjusts the relative magnitude of noise powers of different frequencies in accordance with their relative interference in the reception of speech. This weighting parallels that designed into the telephone instrument.

Based on empirical data, the weighting scheme for different telephone sets may be obtained. A commonly used scheme (for 500 type sets) is the <u>C-message weighting</u>, whose characteristic is shown in Figure 2-20. The significance of this curve is the following: it shows the relative disturbing

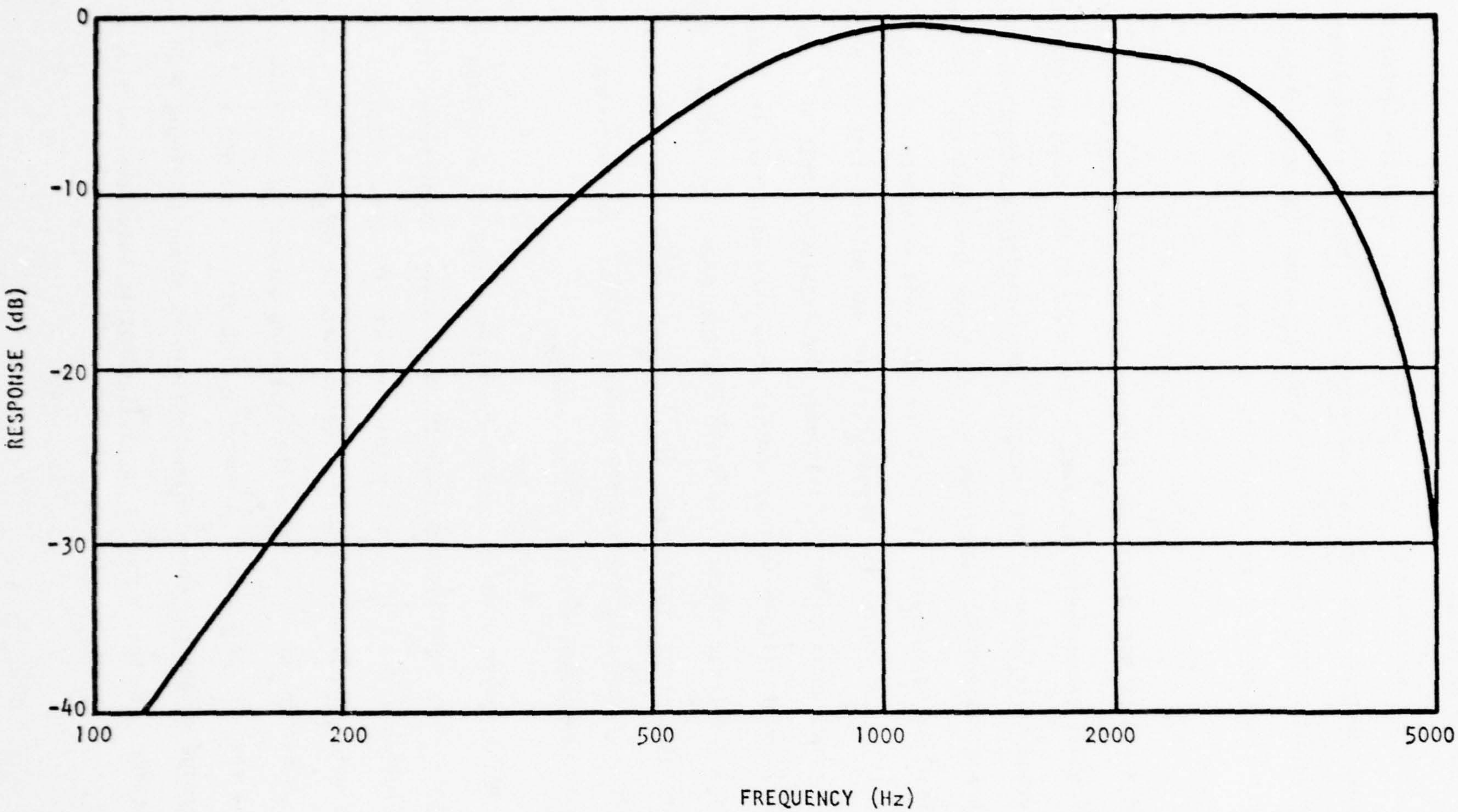


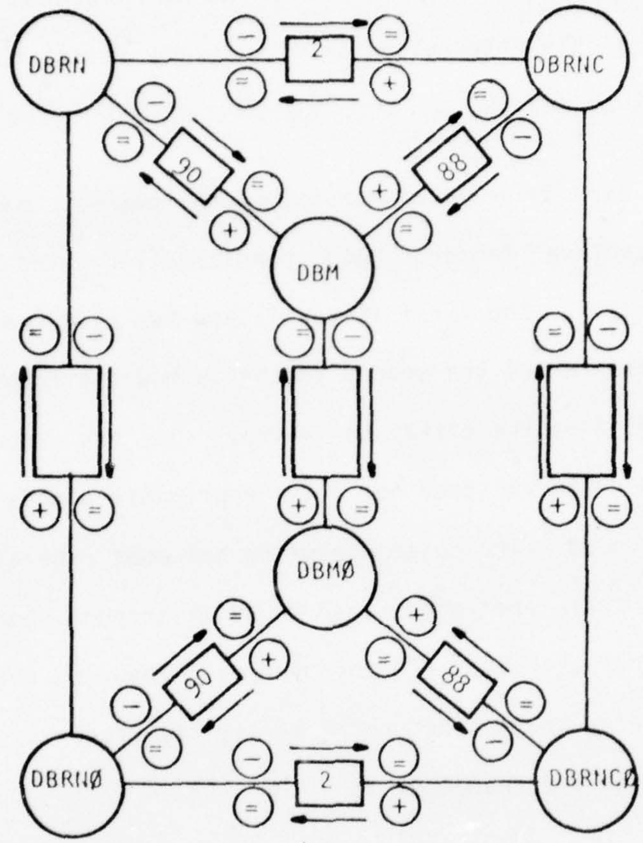
Figure 2-20. C-message frequency weighting

effect of different frequencies. For example, a 200-Hz tone will be 25 dB less disturbing than a 1000-Hz tone. When noise is measured using this scheme, the 200-Hz tone will be attenuated 25 dB more than a 1000-Hz tone by the noise meter; this attenuation is roughly equivalent to that provided by the 500 type telephone set.

b. The commonly used noise meter for telephone channel measurements takes two other subjective factors into consideration which affect an average telephone subscriber. The first factor is how two different noises contribute to the interference; the second factor is how the transient response of the human ear modifies the effect of noise. It has been experimentally found that the human ear does not fully appreciate sounds that are shorter than 200 milliseconds. The noise measuring set must therefore distinguish between sounds that are longer than 200 milliseconds, and those which are shorter. In summary, the noise meter used for measuring noise in telephone channels has three specialized characteristics: frequency weighting, power addition, and transient response.

c. In measuring noise over voice circuits, a noise reference and a scale of measurement is used. This reference is established at 10^{-12} watts or -90 dBm at 1000 Hz. The readings are expressed in decibels above reference noise, abbreviated dBrn. When the C-message weighting network is used, the abbreviation dBrnC is commonly used to indicate the noise level. For other networks, the weighting scheme used should be specified. When the noise readings are referred to 0 dB TLP, the abbreviation dBrn0 is used. When readings in dBrnC are referred to 0dB TLP, the notation dBrnC0 is used. Figure 2-21 provides an easy method for conversion between the various notations discussed.

NOISE POWER LEVEL
CONVERSION DIAGRAM



EXAMPLES USING THE CONVERSION CHART

Set up the equation from left (known values) ro the right (unknown value).

Example 1: DBMØ=?
DBM=-41
TLP=+7

DBM-TLP=DBMØ
-41-(+7)=DBMØ
-48=DBMØ

Example 2: DBRNØ=?
DBRN=4Ø
TLP=-16

DBRN-TLP=DBRNØ
4Ø-(-16)+DBRNØ
56=DBRNØ

Example 3: DBM=?, DBMØ=?, DBRNCØ=5Ø. TLP=+7 (use two steps)

Step 1: Find value of DBMØ

DBRNCØ-88=DBMØ
5Ø-88=DBMØ
-38=DBMØ

Step 2: Find value of DBM

DBMØ+TLP=DBM
-38+(+7)=DBM
-31=DBM

Figure 2.21.

d. The international standard for noise frequency weighting differs from that used by the United States. The International Telegraph and Telephone Consultative Committee (CCITT) has set a different standard for noise measurement and weighting. According to CCITT, noise is measured on a psophometer which has a specified frequency weighting circuit. The term psophometric voltage is used to refer to the rms weighted noise voltage at a point and is usually expressed in millivolts. Since average noise power is used more often than noise voltage in design calculations, the psophometric voltage is converted into an equivalent average noise power delivered to a 600 ohm load. This power is often expressed as picowatts psophometric (pWp). The relationship between psophometric noise voltage and average noise power is:

$$\text{pWp} = \frac{(\text{psophometric mV})^2}{600} \times 10^6 \text{ picowatts} \qquad (2\text{-}28)$$

2-7.7. IMPULSE NOISE MEASUREMENT. The effect of noise on a digital signal is quite different from that on an analog signal. For example, the annoying hiss due to thermal noise in a voice system is of no consequence in a digital system until the amplitude of noise approaches the amplitude of the pulses. However, the effect of spurious impulses appearing in a digital transmission system can be quite consequential. The introduction, or the alteration, of a bit in the bit stream can alter the contents of the entire message when it is received and subsequently decoded.

Impulse counters have been designed to measure impulse noise. Depending upon the amplitude of the pulses being transmitted, a lower amplitude level (threshold) can be set on the counter so that all impulses above the threshold appearing in the system are counted. The distribution of the am-

plitude of the impulses can also be obtained by setting several threshold levels simultaneously. A timer (capable of timing intervals) is also included in the counter to measure impulse noise in a given time interval. In a typical impulse counter for measuring impulse noise in the voiceband, threshold levels adjustable in 1-dB steps from 40 to 90 dBrn with a choice of terminating input impedance are available. The 15- and 30-minute measuring intervals are commonly used for voice circuits.

2-8. EFFECTIVE TRANSMISSION.

The purpose of a telephone circuit is to convey clearly and intelligibly to the destination the message produced by the source. Therefore, to obtain a real picture of the transmission ability of any telephone circuit, it is necessary to take into consideration the loudness or volume received over the circuit as well as the distortion, noise and side-tone, etc. which accompany the message signal. The best measure of telephone transmission is a measure of the circuits' ability to perform this task. Methods of doing this have been developed, and the data thus obtained have been called "effective" transmission data. Some of the factors affecting the quality of transmission are described below.

2-8.1. RECEIVED VOLUME. The fundamental task of a transmission system is to convey the proper signal magnitude at the receiver. The received volume at the destination depends on several things. The conversion of message into an electrical signal by the transducer, the amplification and frequency selectivity of intermediate circuitry, and the attenuation in the transmission medium are some of the factors affecting received volume. In the design and operation of such systems, therefore, proper steps to insure an adequate signal strength at the receiver should be made.

2-8.2. NOISE. Each of the circuits connecting the source to the destination contributes some noise that is transmitted to the receiver at the destination. In telephone transmission systems, a standard limit for the total noise reaching the telephone from all circuits in connection is established; such as 28 dBrnCO for a 60 mile short-haul system and 34 dBrnCO for a 1000 route mile long-haul system. The allowable mean noise for a system of other lengths is then directly proportional to the length. Room noise at the telephone location also contributes noise that effectively adds to the interference received from the circuits. Other sources of noise in a telephone connection are: local loops and trunks.

For transmission purposes, a noisy circuit is equivalent to a quiet circuit whose loss is greater than that of the noisy circuit by the amount of the noise transmission impairment. Even with the normal limits of noise, the actual noise transmission impairment amounts to several dB. However, noise transmission impairment has been taken into account in setting transmission standards and can be ignored unless the noise exceeds the normal limits. For engineering purposes, the noise transmission impairments are merely the additional impairments arising when noise is greater than normal.

Room-noise transmission impairments are very small over a considerable range of room noise, because of the tendency of the listeners in noisy locations to clamp the receivers more tightly to their ears and to cover their transmitters with their hands when listening. Room noise impairments become large only in very noisy locations. In the absence of more concrete methods, it is recommended that the transmission impairment at a telephone in an excessively noisy location be recognized to be 5 to 10 dB worse than at a telephone in a location of normal noise.

In digital transmission systems, the transmission impairment occurs when the amplitude or phase of the signal is changed by noise energy of sufficient magnitude to cause the receiver to make a wrong decision. The normal limit for noise in such systems is specified in terms of signal-to-noise rates for background noise and the number of impulse counts above some threshold over a specific time interval for impulse noise. These limits are based on the data bit error rate acceptable to the user of the system.

2-8.3. DISTORTION. As a signal propagates through the transmission system, its amplitude is attenuated and its phase is shifted. For the signal to be transmitted without distortion, it is necessary that each frequency component be attenuated in the same proportion, and that the phase of the component waves be shifted by amounts directly proportional to their frequencies. Figure 2-22 illustrates the motion of an undistorted signal. The dotted line in A represents a 500-Hz tone with an amplitude of 10 volts. The dot-and-dash line represents a 1500-Hz tone with an amplitude of 8 volts. The solid line represents the resultant input transmitted wave which is composed of 500-Hz and 1500-Hz waves. Figure 2-22B shows the same wave after it is received. Attenuation in amplitude is assumed to be 50 percent. As a result, the 500-Hz wave, shown by the dotted line, now has an amplitude of 5 volts as compared with the 10 volts prior to transmission. Similarly, the 1500-Hz wave is attenuated from 8 volts to 4 volts. The resulting amplitude of the received wave has been attenuated; but note that its shape is unaltered. Note also that the 500-Hz tone, in B, has had its phase shifted by an amount assumed to be 30^{o}. The 1500-Hz tone must have its phase shifted by 90^{o} in order that the phase shifts have the same proportions as frequencies. The phase of the received wave is also shifted, but the shape of the signal remains the same.

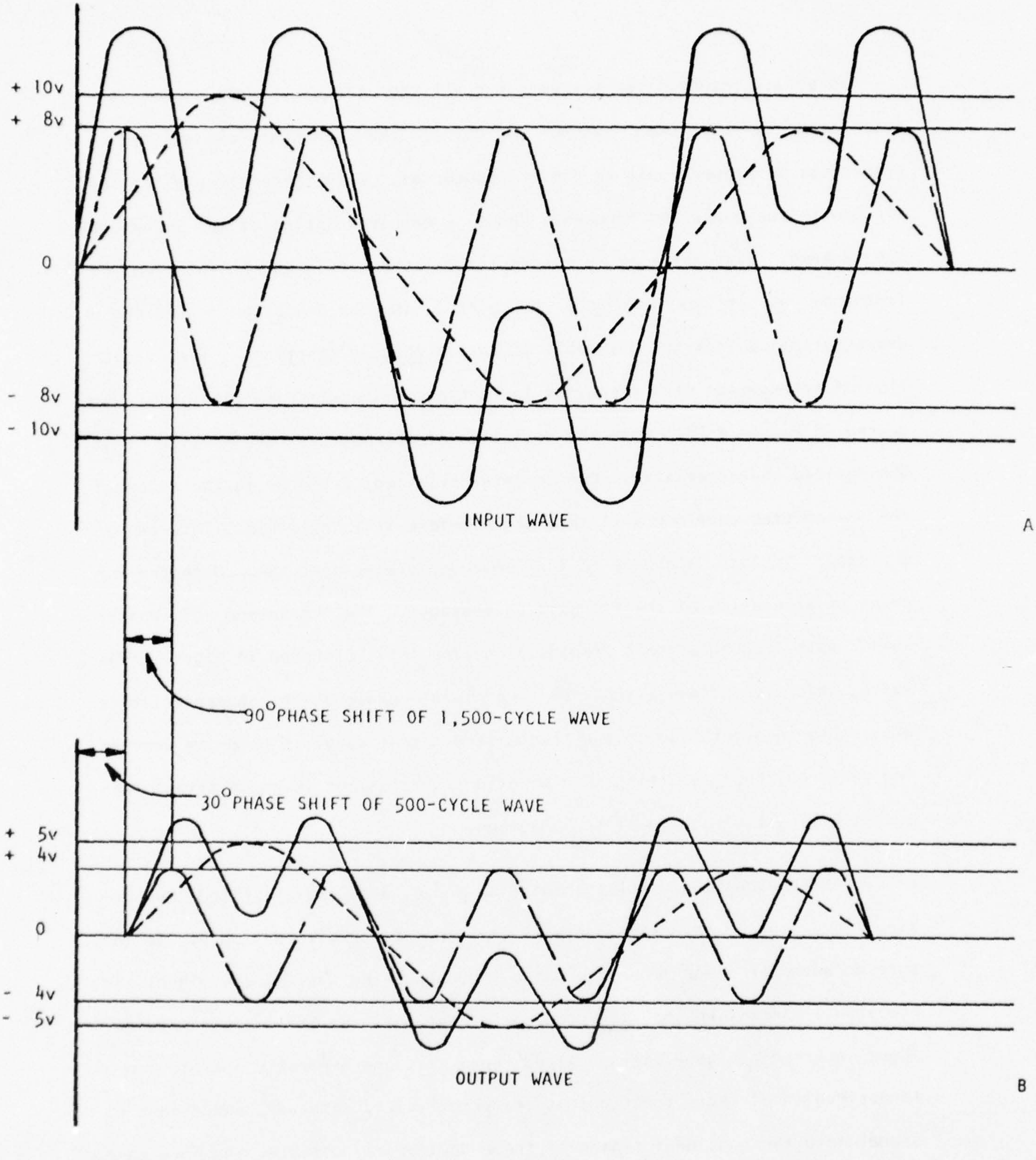


Figure 2-22. Motion of two tones illustrating the desirable transmission conditions.

In actual transmission systems, not all the frequency components are attenuated by the same amount and not all the components are shifted in phase that is proportional to their frequencies. Both these effects are reflected in the phase and frequency response characteristics of the system or sub-systems. The variation in the amplitude attenuation as a function of frequency is called frequency distortion and the deviation in the phase characteristics from the desirable is called phase distortion. The variation of attenuation with frequency for a typical transmission line is illustrated in Figure 2-23. Note the deviation of the actual attenuation from the desired characteristic. Due to this deviation, a change in the shape of the transmitted waveform will take place before it is received. This change in shape implies a change in the component frequencies, and therefore implies an alteration of the transmitted message. The variation of phase-shift with frequency for a transmission line is illustrated in Figure 2-24. Again, this type of deviation from the ideal phase-shift characteristics will cause an alteration in the transmitted signal waveform prior to its being received, thus resulting in distortion. Distortion is an important factor affecting the quality of transmission.

2-8.4. INTERMODULATION. All transmission media are characterized by attenuation, the progressive decrease in signal power with distance. To insure an adequate signal-to-noise ratio throughout the system, the signal is repeatedly amplified as it propagates through the transmission medium. These intermediate amplifiers, called repeaters, are generally operated in a linear region of their input-output characteristics. However, excursions of signal into the nonlinear region of these devices is possible, and does take place quite often. When this happens, frequencies resulting from the sums and differences of the various frequency components of the message signal

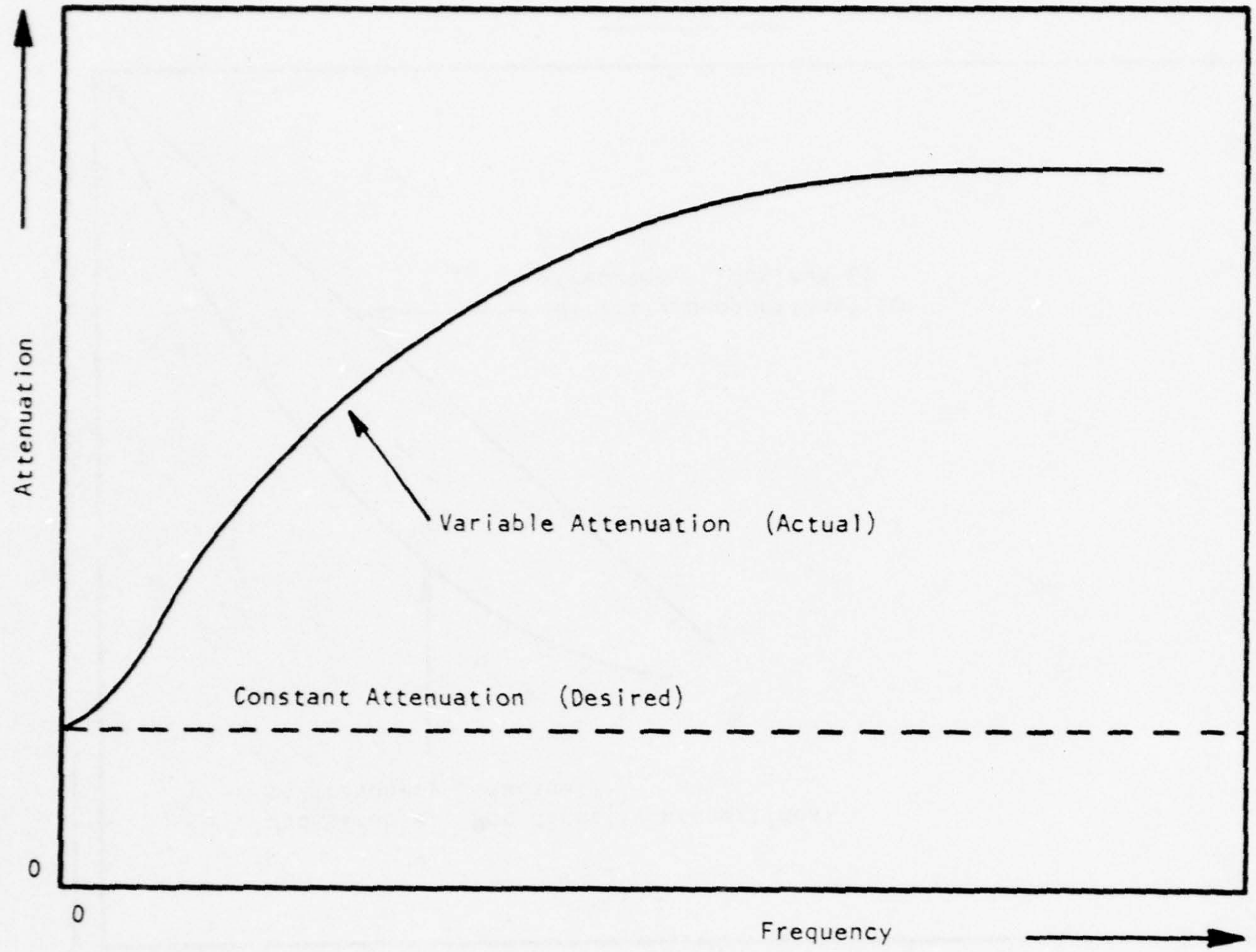


Figure 2-23. Variation of amplitude attenuation with frequency.

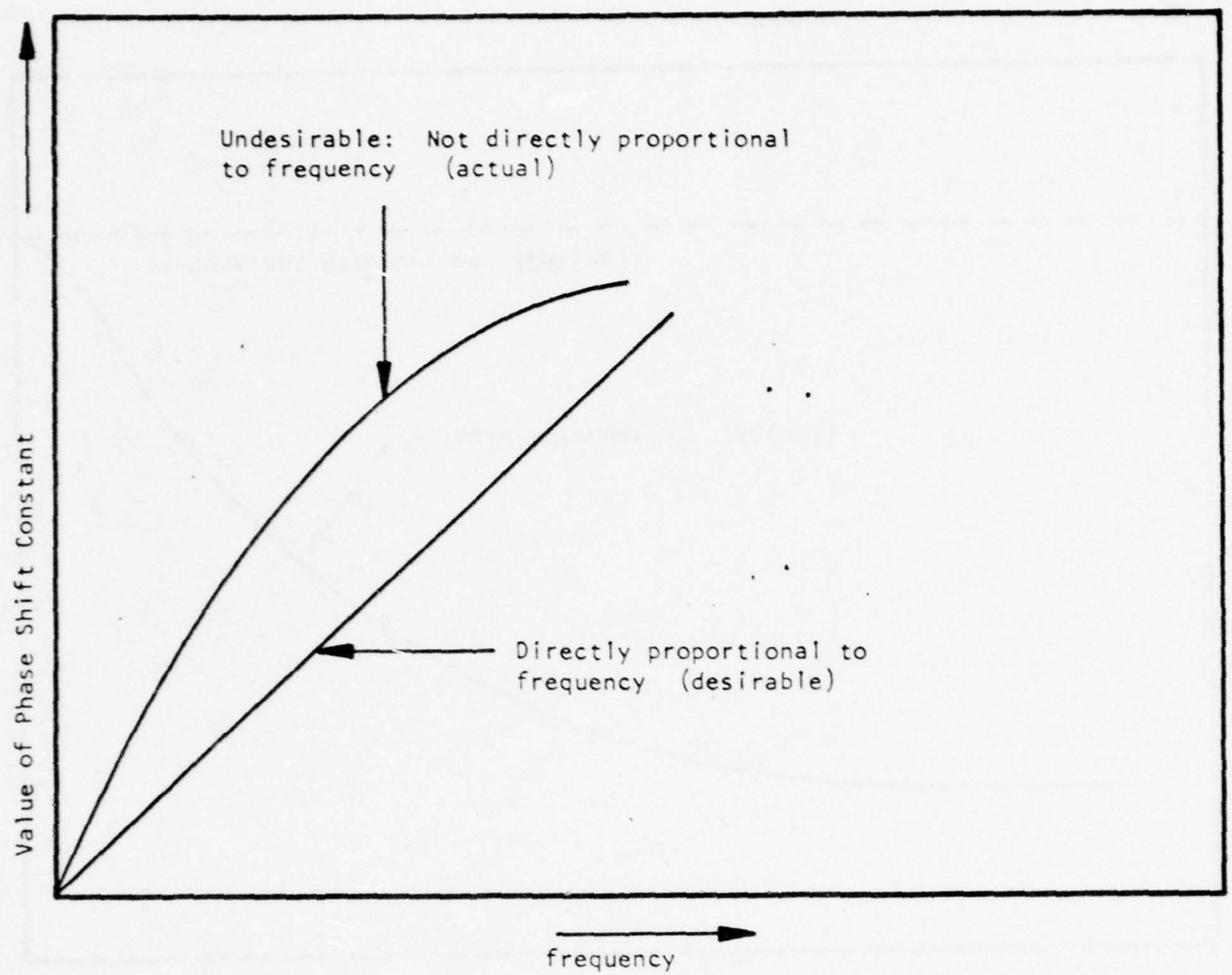


Figure 2-24. Variation of Phase-Shift Constant with Frequency

are generated. This phenomenon is termed intermodulation. The intermodulation effect is undesirable and can be quite disturbing.

2-8.5. ECHOES. The term echo, as its name might imply, refers to the reflection of a wave from points in the transmission medium. This reflected energy, through amplification, may reach a certain point with sufficient magnitude and time delay such that it cannot be distinguished from the directly transmitted wave. This effect can seriously impair the data transmission capabilities of a given channel by causing excessive error rates if the amplitude and time delay of the reflected signal are significant with respect to the transmitted signal. In voice transmission, echoes can be quite annoying to the telephone users. The principal causes of echoes in wire-line circuits are impedance mismatches in the system. In order to remove echoes, the reflected signals are severely attenuated by the use of what are called echo suppressors.

2-8.6. CROSSTALK. Originally any undesired speech heard on a circuit caused by speech over another telephone circuit was called crosstalk. The meaning of the term has been broadened to include undesired signals in any communication channel that originates in another associated channel. Crosstalk is considered to be a transmission loss and is undesirable.

Parts of complete communication circuits that are physically close together are usually coupled by magnetic and electrostatic induction to some extent. This causes signals in one circuit to produce undesired signals in another circuit. Such coupling could, theoretically be eliminated by complete electrical shielding, but this is usually impractical. In a multiconductor cable the pairs are twisted with respect to one another so that the alternate positive-negative coupling leaves a net of near zero. When sys-

tematic relative twisting is impractical, such as when a new line is run in near an old one, partial shielding and as much physical separation as possible is good practice.

SECTION III

FUNDAMENTALS OF TELEPHONE TRANSMISSION

3-1. GENERAL

The transmission of speech over telephone facilities is the most widely known form of transmission. Voice frequency signals are transmitted over the nominal 4 Hz voice channel. Through multiplexing techniques many voice channels are combined to form wideband signals for transmission over the long-distance network. These wideband facilities also handle wideband signals such as television video. Narrowband signals, such as telegraph, may also be combined together for transmission over a single 4KHz voice channel. Thus, many of the factors affecting the generation, transmission and reception of voice-frequency signals over the existing telephone facilities also have a bearing on the transmission of other signals over the same facilities.

This section describes the generation, transmission and reception of telephone voice-frequency signals, the basic limitations and problems encountered in telephone transmission, and the solutions to some of these problems. The transmission of discrete-time signals (digital) over analog (telephone) facilities is also introduced.

3-2. TELEPHONE INSTRUMENT

The telephone set provides the interface for transmitting and receiving voice, and can provide the interface for transmitting and receiving data over a telephone network. The telephone set is composed of a transmitter, a receiver, and additional circuitry for equalization and signaling.

3-2.1. TRANSMITTER. The transmitter is essentially a transducer that converts the acoustic energy into an electric signal. It is composed of two electrodes: one electrode is fixed while the other one is movable. In most of the present day telephone sets, the two electrodes are separated by carbon granules. The resistance between the electrodes due to the carbon granules can be varied by changing the position of the movable electrode. When the handset is removed from the cradle or hook, a dc potential is applied to the electrodes, resulting in a dc current between them. During conversation, the acoustic waves collide with the movable electrode; the vibration of the electrode causes the resistance between the electrodes to vary, thus producing a current which is varying in amplitude in accordance with the impressed acoustic waves (modulated current).

3-2.2. RECEIVER. The receiver is again a transducer, and performs the inverse function of the transmitter. The modulated current received is passed through a winding mounted on a permanent magnet. The varying current produces a varying magnetic field; the alternate increase and decrease in the magnetic field causes a diaphragm, made of a magnetic material, to move. This movement of the diaphragm produces acoustic waves similar to those produced by the talker at the distant telephone. The transmitter and receiver efficiencies are measured, respectively, in terms of how well the acoustic energy is converted into the electric signal, and vice versa. These efficiencies are inter-dependent and have a significant bearing on the overall quality of transmission.

3-2.3. SIDE-TONE. The fact that the receiver and transmitter in the handset are connected to the same pair of wires causes the talker to hear

his own voice through his own receiver. Such "feedback" phenomenon is called sidetone. Clearly, the amount of sidetone must be controlled, but not eliminated, since empirical tests indicate that some amount of sidetone is essential to user satisfaction during a telephone conversation. When the sidetone level is high, the normal reaction of the talker is to lower his volume, thus reducing the output level of the transmitter. If the sidetone level is low, the conversation sounds unnatural, and people tend to talk loudly. Anti-sidetone circuitry is incorporated in the telephone set to control the amount of sidetone introduced and the dc path between the transmitter and the receiver is always blocked by using capacitive coupling.

3-2.4. DIALING. The dialing signal to the central office is also generated by the telephone set. Two types are in current use; DC and tone. The DC dial pulses are generated within the set by a switching mechanism. The dc battery current which flows from the local central office to the subscriber's telephone set during the off-hook condition is interrupted by the closing and opening of a switch in response to dialing. Thus, pulses are formed as each digit is dialed by the subscriber. These dial pulses are used for address and control information at the central office. The tone dialing signals are generated from a group of audible frequencies; a pair of frequencies is generated to represent each dialed digit. This form of address signaling is called dual-tone multi-frequency (DTMF) dialing.

3-2.5. 500 D-TYPE TELEPHONE. The telephone sets designed in recent years--e.g., 500-type telephone sets--have additional circuitry incorporated in them to improve the frequency response of both the transmitter and the receiver, and also increase the receiving and transmitting efficiencies.

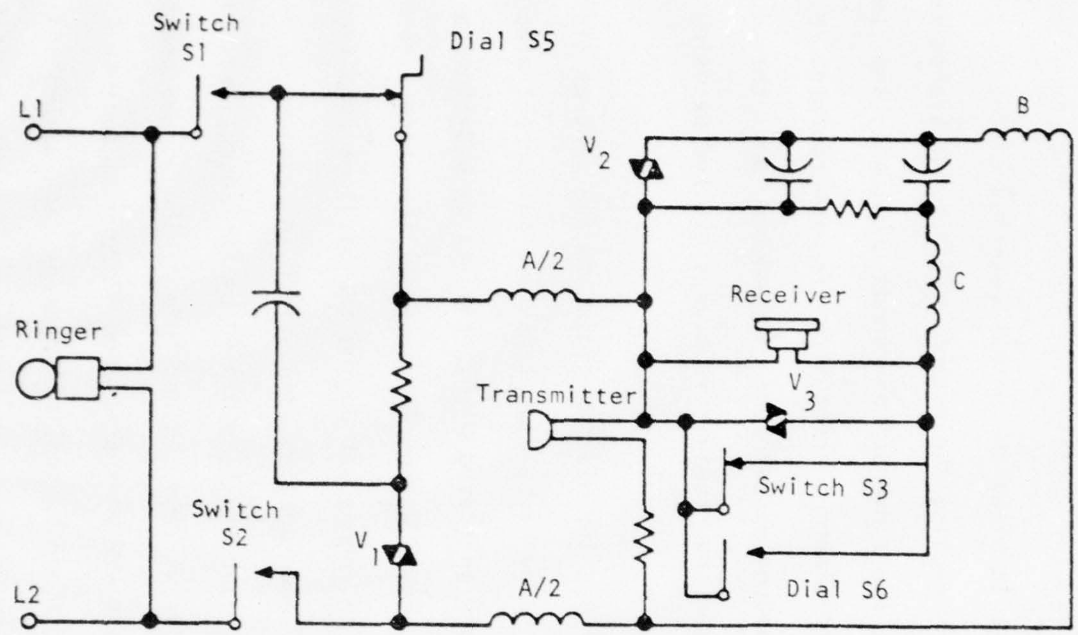


Figure 3-1. Schematic diagram of 500D-type telephone set.

Equalization circuitry, comprised of basic R, L, C elements and nonlinear resistors, is used in these sets to help make the speech volume at the central office and at the customer's receiver less dependent on the wire-pair length. The nonlinear resistors (called varistors) in the equalizing network also compensate for differences in customer wire-pair impedances (due to varying lengths) which would otherwise tend to produce unbalance in the sidetone circuit. A schematic diagram of the 500 D-type telephone set is shown in Figure 3-1. During the on-hook condition, switches S1 and S2 are open; switch S3, which controls the flow of the ringing currents across the receiver and the transmitter, is closed; the voltage across the varistor V3 is therefore zero, thereby protecting the transmitter and the receiver from ringing currents. When the handset is removed, switches S1 and S2 are closed. Direct current from the central office on lines L1 and L2 can now flow through the transmitter, and open the switch S3, thereby removing the short-circuit across the receiver. During dialing, the dial contact S5 opens and closes to form the dc pulses for each dialed digit. These pulses are transmitted to the central office for routing and control purposes. Also, switch S6 is closed during dialing, thereby disengaging the receiver. Varistor V3 helps to suppress clicks (pulses) in telephone receivers, while varistors V1 and V2 perform equalization functions described above.

3-3. TELEPHONE AND INTER-OFFICE CONNECTIONS

The telephone sets of the calling and the called parties are interconnected via a switch or network of switches. The pair of wires connecting the subscriber's telephone to the base dial central office (DCO) is called the subscriber loop. The DCO is connected to the commercial network by inter-office trunks. The transmission objectives, defined in terms of al-

lowable losses, interference, received volume, etc., for subscriber loops and inter-office trunks are allocated separately.

3-3.1. SUBSCRIBER LOOP. The pair of wires connecting the subscriber's telephone to the DCO forms the subscriber loop. The voice and signaling information between the subscriber and the DCO is carried by these wires. Thus, the dc current for the telephone transmitter during off-hook condition, the ac ringing signal, the dial pulses or DTMF for address and control, and the voice signal are all transmitted over the subscriber loop. Clearly, the proper design of the subscriber loop plant is of great importance. Satisfactory service requires that the loop plant be such that the transmission loss does not exceed the limiting loop loss (which is the maximum loop loss permitted in the given DCO area) and that the dc loop resistance does not exceed the limits of the central office equipment.

The selection of wire sizes and loop lengths is basic to the design of subscriber loops. This selection is directly affected by the relationship between the resistance and attenuation, the limitations placed on supervisory signaling and transmission by the type of DCO equipment, and the effect of changes in ambient temperature and humidity conditions. The basic criteria used in establishing the wire sizes and loop lengths are the following:

a. Attenuation limits: This criterion falls under the category of transmission design, where the attenuation of the message signal is considered. A reference frequency of 1000 Hz is used in the United States to set the standards on the attenuation limit.

b. Signaling Limits: This criterion falls under the category of resistance design, where the dc voltage drop (=IR) of the line is an impor-

tant consideration. Most signals used in supervision and control are dc. Thus, the IR drop, which is directly proportional to the loop resistance and concomitantly proportional to the loop length, is of consequence. If the length of the loop is extended, the resistance R also increases. A point is reached where the IR drop is so high that the signaling and/or supervision is no longer effective. With modern telephone sets, supervision is usually the limiting factor. When a subscriber goes off-hook, the telephone loop is closed and a dc current flows. This current is used to trigger a relay at a switch at the central office. If the loop length is unduly long, the amount of current flow would be insufficient to trigger the relay in the desired manner, thus making the supervision ineffective. However, where such long loops are unavoidable, loop-extender circuitry can be added (dial-pulse repeaters).

Since the percentage of time a customer loop is used by the subscriber is small, line concentration techniques are often used between the customer and the DCO. Thus, instead of having a pair of wires between each telephone set and the DCO a concentrator is used to connect the subscriber to the DCO; this allows many subscribers to <u>share</u> a limited number of lines to the DCO. Due to this sharing of lines, the number of lines needed between the concentrator and the DCO is smaller than the number of subscribers connected to the concentrator. This type of concentrator is known as a private branch exchange (PBX) or as an automatic PBX (PABX).

An important factor that effects the design of a subscriber loop is the type of telephone sets used. Thus, design considerations involved in subscriber loop design using 500-type sets are different than those for other type sets.

3-3.2. INTER-OFFICE CONNECTION. The DCO is connected to commercial exchanges by means of inter-office trunks. The interoffice trunk might be made up of one or more wire pairs carrying a group of multiplexed channels or it could be a larger number of individual pairs in a cable (one pair for each trunk).

On commercial telephone calls, several types of trunks might be involved in a typical call completion. An interoffice trunk connects the DCO to a commerical end office; another interoffice trunk could connect the commercial end office to another commercial end office; a tandem trunk connects an end office to a tandem office; and a toll-connecting trunk connects an end office or tandem office to a toll office. Like the interoffice trunk, the tandem and toll-connecting trunks may be one channel of a multiplexed channel group or an individual channel. Intertoll trunks interconnect toll offices. These are almost exclusively multiplexed channels over wideband transmission media. Figure 3-2 is an illustration of how these various types of trunks may be involved in a telephone connection. It should be noted that the symbols used in Figure 3-2 are standard symbols for various offices- and trunk-types. In large areas, shown as area A in Figure 3-2, several end offices may be present. An additional exchange, called a "tandem office," is used for switching lines between the various central offices. In inter-city customer-to-customer connections, a toll-office is usually involved (as depicted in Figure 3-2). Thus, all calls from area A to area E go through the toll-offices, using the toll-connecting trunks. The primary- and sectional-centers shown in Figure 3-2 are part of the hierarchy of switching centers used in toll-switching plan on commercial calls.

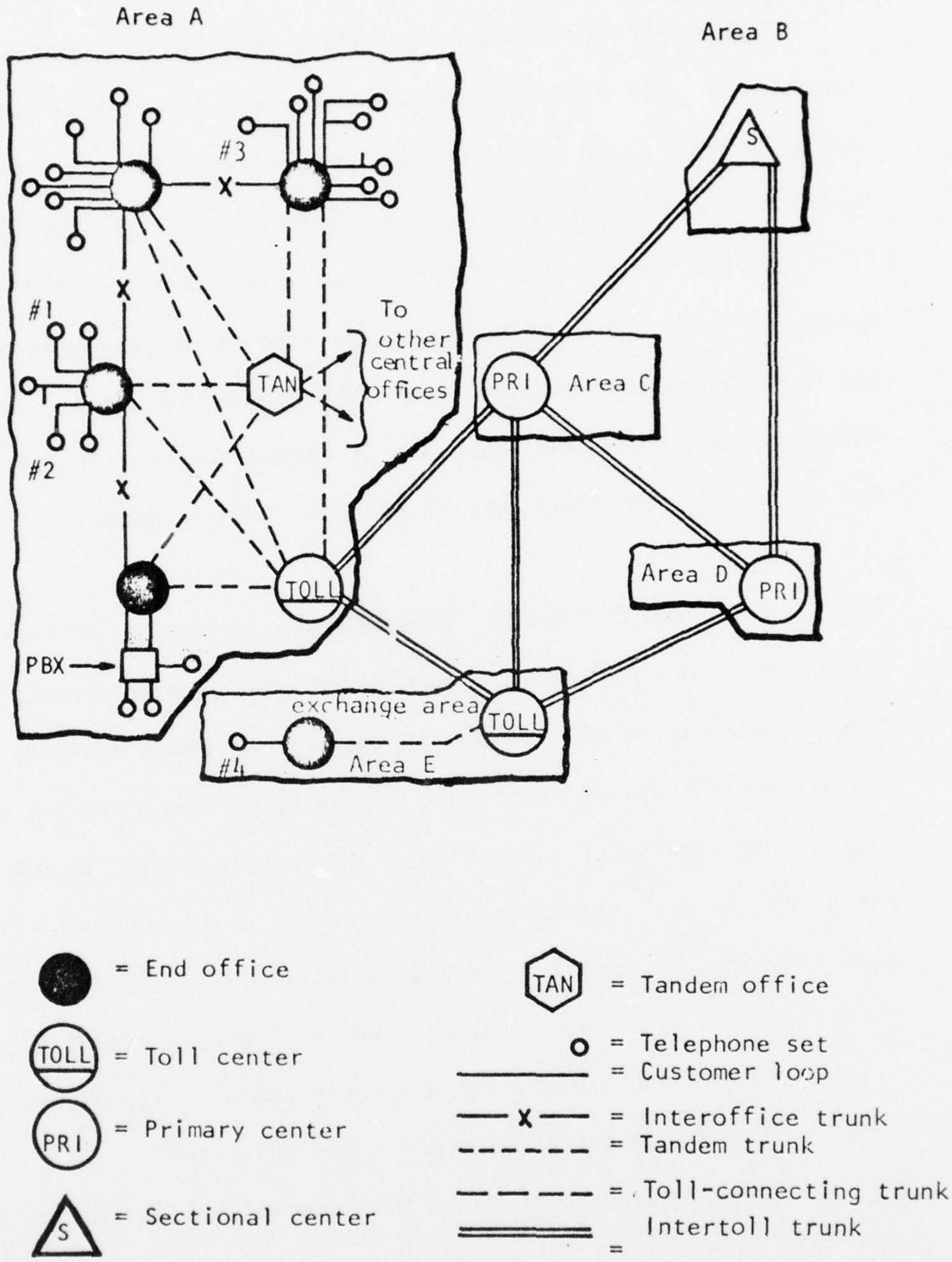


Figure 3-2. A simplified telephone network

A trunk circuit includes the office equipment on both ends, and the transmission media between the two offices. In order to meet the transmission objectives, loading of trunks (which involves the insertion of series inductance periodically along the trunk to compensate for frequency distortion) and the use of amplifiers (repeaters) are usually required. The design of the DCO and local exchange area plant are extremely important to the fidelity of transmission over the entire network. Such unwarranted phenomena as singing and echo (to be discussed later in this section) arise due to impedance discontinuities and poor return losses.

3-3.3. SWITCHING OFFICES IN TRANSMISSION. Switching offices or exchanges encountered in telephone transmission can be either manual or automatic. Manual switching is performed with the help of human operators (switchboard) and use of this form of switching is mostly for establishing special connection and international connections. In most existing automatic switching, banks of relays and mechanical switches are used to perform the various functions of a switching center; in more recent switches, computers are used to control the switching functions; and in the most recent switches, connections are established without electro-mechanical devices, being accomplished through a technique called time divisions switching.

The basic stages involved in the processing of a typical call are similar in each of the switch types, and consequently, the basic functions performed by a switching center can be discussed in generic terms. Before embarking on a discussion of these functions, however, the following discussion of the different types of switches is presented.

a. The types of electro-mechanical switches commonly found are the step-

by-step Strowger switches and cross-bar switches. In step-by-step switches, banks of wafers (contacts) are arranged in a two-dimensional array. A moving wiper, capable of vertical and horizontal movement in response to the received pulses, establishes a contact between the input line and an output line whenever service is desired and available. In some step-by-step switches, each incoming pulse moves a counter which stores the digits in a register. The input to the step-by-step switches is then fed from this register as it is needed. Cross-bar switches are more recent than step-by-step switches. Like other switches, cross-bar switches connect the input wire to one of a number of possible output wires. Cross-bar switches, as the name implies, essentially consist of sets of horizontal and vertical bars which can be moved by electro-magnets. This movement is such that it causes a relay-like contact to be made at the coordinate intersection of the bars. By a suitable arrangement of these intersections in response to the received pulses, a path can be selected through the exchange. Cross-bar switches are faster than step-by-step Strowger switches, and less prone to generating impulse noise. Presently, there are some electronic (computer controlled) cross-bar exchanges.

b. A computer-controlled switching exchange basically comprises an electro-mechanical switching mechanism, and a computer system. The subscriber lines and trunks enter the switching mechanism in which the interconnection paths can be set up. The switches used may be cross-bar or reed switches, called ferreeds. The switching mechanism is organized into frames, and each frame has a controller, called the switching controller, to control the set up of line paths in that particular frame. The opening and closing of the switches in the switching mechanism is controlled by the com-

puter program. Two types of memory units are used by the computer: semipermanent memory which contains programs and fixed data about the lines, and is a "read only" type unit (thus, avoiding any overwriting of the program during normal operations), and temporary memory which contains information about the calls in process and other information which changes during normal operation. The information regarding status of the lines, trunks, and signal receivers (for signaling) is carried out by line scanners, and is continuously fed to the computer for appropriate processing.

c. In a time-division multiplex switch (TDMS), the switching control is accomplished by a computer in a manner similar to that discussed above. However, the switching mechanism is electronic rather than electro-mechanical. Connections are actually established by the manipulation of "time-slots" between the incoming and outgoing sections of the switch mechanism. The basic operation is similar to Time Division Multiplexing as discussed in section 6.5 below.

3-3.4. FUNCTIONS OF A SWITCHING OFFICE. The basic stages in the processing of a typical call are as follows:

a. When the subscriber goes off-hook, the office must detect that service is needed. Dial tone is connected at that time (indicating that the subscriber is connected to a dialing receiver) and the switch waits for the subscriber to dial.

b. The dialed telephone number is received and used to set up an interconnection.

c. The number of incoming lines at an office is usually larger than the number of trunks. The switch therefore concentrates the calls into a limited number of paths through the exchange; if no path is available, a

trunk-busy signal is provided to the subscriber.

d. The path through the exchange is then connected to the desired called party line, either directly if it is in the same DCO, or via the switching hierarchy if it is in another office; if the desired called-party line is busy, the switch notifies the user by means of the busy signal.

e. If a complete path for the call is found, the "terminating" exchange makes the telephone of the called person ring.

f. When the called party answers the telephone, the ringing signal is removed.

g. Once the call is successfully established and completed, the circuit is disconnected when the calling party goes on-hook.

h. Finally, the exchange also keeps track of the charges for each customer.

3-4. TELEPHONE CHANNELS

The telephone set and the local area equipment and the facilities to which they connect are usually two-wire circuits. These circuits use a single pair of wires to carry the message signals in both directions. In the trunking network, four-wire circuits are employed which use a separate pair of wires for each direction of transmission. The economic advantage of using two-wire circuits would seem to dictate their use. However, the trunking network is mostly composed of multiplexed wideband facilities which by design, provide four-wire circuits. There are, however, some advantages with four-wire circuits for special circuits in the local area.

3-4.1. TWO-WIRE CIRCUITS. In order to meet the transmission objectives for an interoffice or tandem trunk, it is often necessary to add gain to the

transmission path; amplifiers (repeaters) are used to provide this required gain. A number of special problems are encountered when amplifiers are used in a two-wire system: the gain introduced must be independent of the direction of transmission; and dial pulses for signaling, etc. must not be impaired. A repeater used in typical two-wire paths is shown in Figure 3-3. Two one-way amplifiers are connected by hybrid coils between the two lines to provide independent amplification in each direction of transmission. Considering transmission from line A to line B, half the power from line A(P_S) is transferred to the upper amplifier (A_1) by hybrid coil No. 1. The other half enters the output of the lower amplifier and is lost. The amplified power, P_A, from the upper amplifier is delivered by hybrid coil No. 2, half (P_R) to line B and the other half to balancing network B where this half is lost. Since each hybrid coil divides the power in half, there is a 3 dB loss in each, and the net gain of the repeater is therefore 6 dB less than the gain of the amplifier alone. Ideally, the impedance Z_{NB} of the balancing network B should exactly equal the impedance Z_B of line B; for if this were true none of the power from amplifier A_1 would be transmitted across hybrid coil No. 2 to the input of amplifier A_2 for further amplification and transmission back into line A. In practice, this perfect balance is unattainable. Network B is made to balance closely the nominal characteristic impedance of line B, but the actual impedance of the line may differ from its nominal characteristic impedanace because of irregularities in the line, or imperfect termination at the far end. Therefore, there is usually some degree of unbalance that causes a small part (P_U) of the signal to reach the lower amplifier and to be returned like an echo through the lower amplifier to line A. <u>Return loss</u> is a measure of the degree of unbalance.

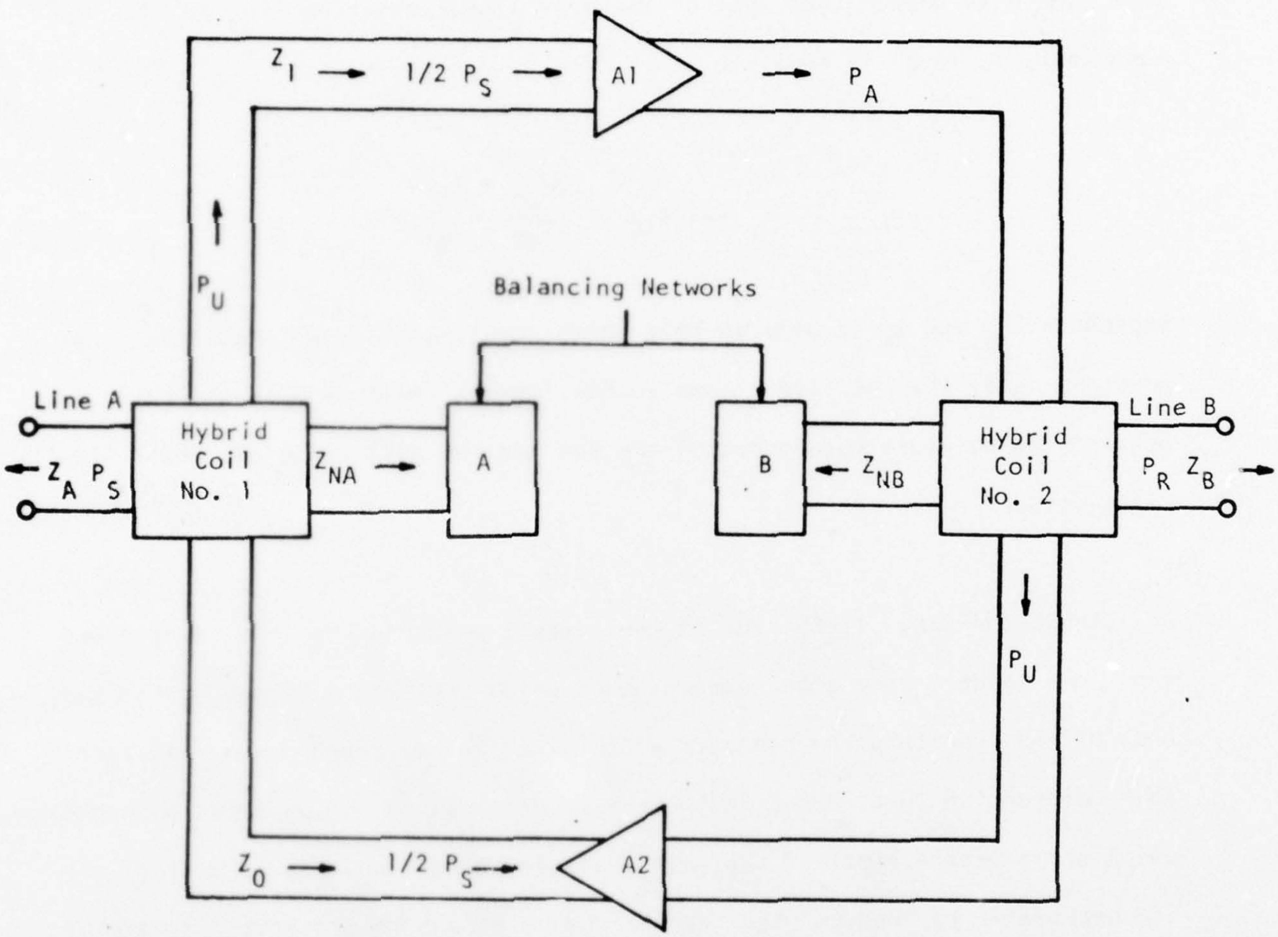


Figure 3-3. Repeater Used in Two-Wire System

a. RETURN LOSS. The unbalance between network B and line B, Figure 3-3, behaves just the same as though a reflection occured at the input of line B, sending part of the signal back through the repeater to line A. The return loss is the number of dB that the power of this fictitious reflected signal from line B is weaker than that of the main signal entering line B. The return loss, ρ, in dB is equal to

$$= 20\log_{10} \frac{|Z_{NB} + Z_B|}{|Z_{NB} - Z_B|}$$

Impedance Z_{NB} and Z_B in general have both resistance and reactance components, and the vertical bars in the formula indicate that the ratio is between the absolute magnitudes of the sum and the difference of these two impedances.

b. SINGING MARGIN. If the sum of the overall gains in the two directions from one line to the other exceeds the sum of the return losses at the two ends of the repeater, the repeater will "sing" at any frequency. Considerable distortion will occur and make the speech sound hollow, even when the gains are a little short of the actual singing condition. In practice, it is necessary to reduce the repeater gains enough to preserve a margin of safety against the occurence of this condition. A common rule is to require a singing margin of 8 dB; that is, to require that the sum of the gains in the two directions (between the line terminals of the repeater) be at least 8 dB less than the sum of the return losses. Thus, the gain obtainable in a two-wire repeater is definitely limited by the degree of balance between the networks and their lines.

c. SINGING PATHS AND STABILITY. As the length of a system increases and more repeaters are used, the increase in the number of possible singing paths is greater than the increase in repeaters. For example, in a 3-repeater system there are, in addition to the three separate paths around each repeater, two paths each involving two repeaters and one path involving all three repeaters, or a total of six paths. A safe margin against singing must be maintained in all of these paths, separately and in combination. Because of the increased number of singing paths, the repeater gains are more limited and the circuit net losses are greater on long circuits having a number of repeater sections.

d. CROSSTALK. Crosstalk, especially that induced at the near-end of a repeater limits the gain that can be used in each repeater and increases the net loss of the system. Near-end crosstalk is accentuated by the repeaters in a line since is is amplified by the repeater.

e. NEGATIVE IMPEDANCE REPEATER. The negative impedance repeater is a device which acts like a negative impedance connected in series with a line and provides gain for signals traveling in either direction of a two-wire path. Repeaters of this type preserve the d-c continuity of the circuit and therefore, transmission, although degraded, is not interrupted upon repeater component failure.

3-4.2. HYBRID TWO-WIRE/FOUR-WIRE CIRCUITS. A circuit such as shown in Figure 3-4 affords better performance than the two-wire system since there is only one possible singing path, the one-way transmission channels can be regulated to hold the overall attenuation nearly constant, and near-end

cross talk is reduced by using two pairs which are separated in a single cable or contained in separate cables. This type of circuit is typical of those in the telephone network at points where the local area two-wire system interfaces with the four-wire trunking system.

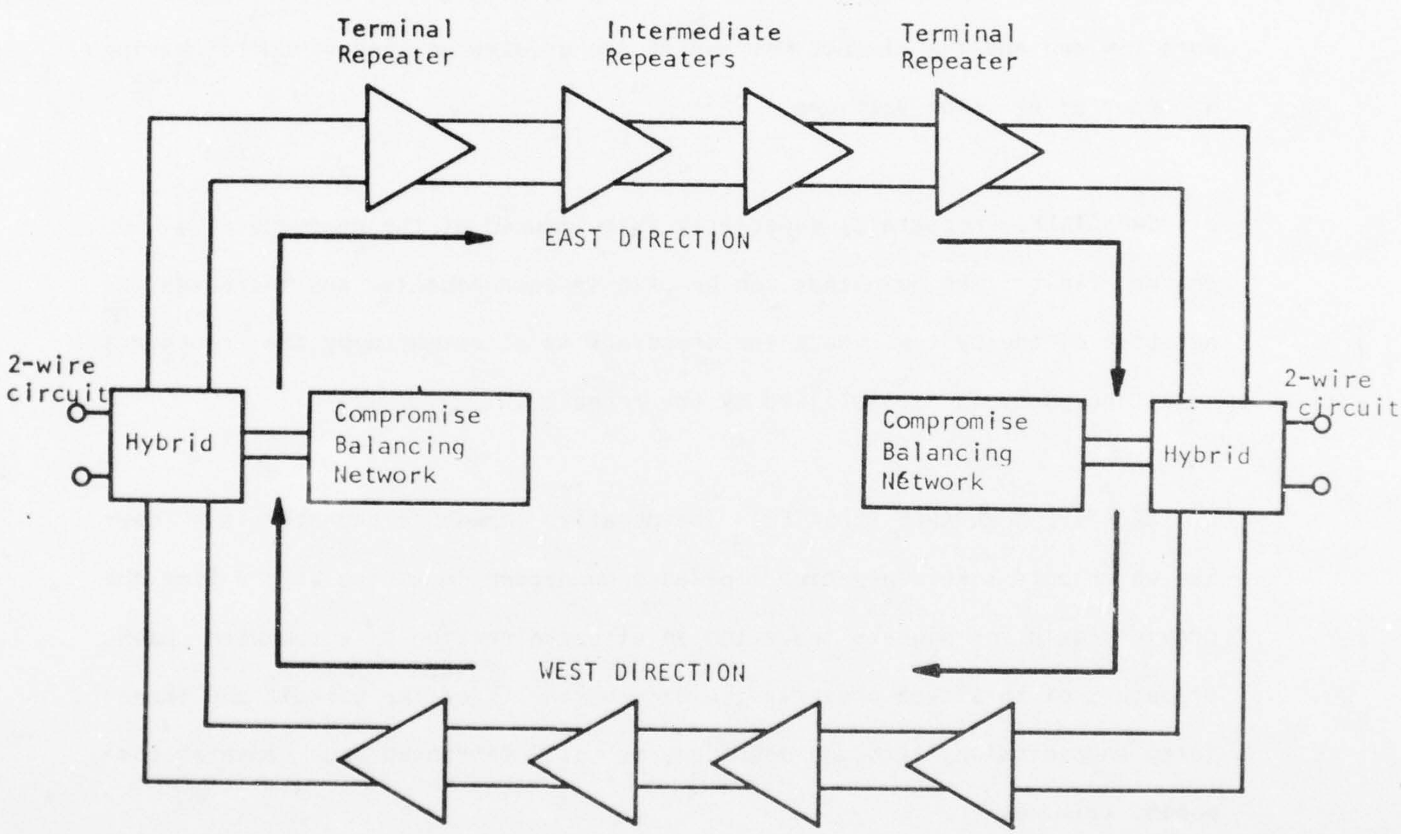


Figure 3.4. Hybrid Two-wire/Four-wire Circuit

a. NET LOSS LIMITATIONS. The sum of loss at the two ends of the circuit must be at least 8 dB at the worst frequency. The hybrid circuit at the ends of the circuit, Figure 3-4, are usually referred to as four-wire terminating sets with compromise balance networks. Although these networks provide return losses as low as 4 dB, four-wire systems operate with adequate singing margins at net loss as low as 0 dB regardless of the length. However, when the circuit is long, the presence of echo increases the net loss.

b. REPEATER SPACING. The spacing of four-wire repeaters is determined by noise and cross talk. One requirement is that the line sections between repeaters must be kept sufficiently short so that the level of the signals is never too close to the level of the interference. On the other hand, each repeater section makes the same contribution to the interference (in terms of power) that appears at the end of the circuit, so that the greater the number of sections of a given length, the larger is the interference at the terminal. The best repeater spacing is a compromise between these two factors.

c. ECHOES ON TWO-WIRE/FOUR-WIRE CIRCUITS. Speech waves traveling down a transmission line are reflected back toward the source if there is a mismatching termination. In the case of a two-wire circuit, the relation between the strengths of the wave reflected back from the terminal and the wave that arrived there is measured by the return loss at the distant end of the circuit. If the line is short, the reflected current may return so quickly that it is indistinguishable (to the human ear) from the transmitted current. If the loss is high, the returned current may be too weak to hear.

In either case, the returned current is no problem. In the hybrid two-wire/four-wire path, the return current is delayed by twice the time of propagation of the circuit--the time for the speech to travel to the far end plus the time for the reflected speech wave to return. In loaded cable pairs, the speech signals travel with a velocity of about 15 miles per millisecond. Therefore, in a 500-mile circuit, the reflected current would return to the talker after 66 milliseconds (ms). This delay is sufficiently large for the returned current to be heard as a distinct echo and these are disturbing to talkers. Loud echoes are naturally worse than weak ones and echoes having long delay are worse than those having short delays. One way to control echoes is to adjust the overall loss of the circuit so that the echo is weak enough to be tolerable. Thus, there is a minimum net loss at which a circuit can be satisfactorily operated because of echoes. The greater the delay of the circuit, the larger the minimum net loss.

3-4.3. FOUR-WIRE CIRCUITS. Although four-wire circuits are the rule in the trunking system, they are not used in the common telephone subscriber loop plant. Four-wire circuits are used on some special high-quality telephone systems, and on many data circuits which utilize telephone channels. The advantages of an end-to-end four-wire circuit are, of course, the absence of the echo and singing problems introduced by the hybrid coils in the two-wire and two-wire/four-wire circuits.

3-5. DEFENSE COMMUNICATION SYSTEM (DCS) AUTOVON. The DCS includes DoD world-wide, long-haul Government-owned and/or leased, point-to-point circuits, trunk terminals, switching centers, control facilities and tributaries required to satisfy the communications needs of Department of Defense

AD-A065 744 PURDUE UNIV LAFAYETTE IND SCHOOL OF ELECTRICAL ENGI--ETC F/G 17/2
ENGINEERING FUNDAMENTALS FOR BASE WIRE TRANSMISSION.(U)
OCT 78 H K THAPAR, B J LEON, T H WEAVER F30602-75-C-0082
UNCLASSIFIED 1842 EEG-EEIC-TR-79-7 NL

2 OF 3
AD
A065744

Agencies, Commands, headquarters and other subscribers, both military and civilian. In serving its users, the DCS handles long- and short-haul communications traffic of various priorities and functional categories. This includes command and control, operational, administrative, logistical, weather and intelligence traffic. Traffic over the DCS may be in the form of voice, teletype, graphics and data. The DCS employs a combination of commercial and military communication facilities, organized on a zonal basis and connected by inter-zonal transmission lines. The present DCS has three common user networks. One is the Automatic Voice Network (AUTOVON); this network is based on the nominal 4-KHz bandwidth voice channel, and is capable of handling voice and other traffic which can be coded into waveforms compatible with its channel characteristics. AUTOVON has a precedence arrangement that gives high priority users preemption capability. A second is the Automatic Digital Network (AUTODIN); this network sends alphanumeric traffic between users at rates from 75 b/s to 4.8 kb/sec. User access to the AUTODIN is at various bit rates related by $2^n \times 75$ with $n = 0, 1, \ldots, 6$. The third network is the Automatic Secure Voice Network (AUTOSEVOCOM). This network transmits encrypted voice at three basic rates: 2.4 kb/sec, 9.6 kb/sec, and 50 kb/sec.

In this section, the AUTOVON is discussed.

a. All AUTOVON access lines to an AUTOVON switch are four-wire. Two-wire users have the capability of connecting into AUTOVON, but only through another switch, such as the base DCO or a PABX facility. The interconnection between the DCO/PABX and the AUTOVON switch is either a four-wire circuit or is converted to a four-wire circuit by means of the four-wire termi-

nation set (hybrid). The types of interconnections permitted are shown in Table 3-1. All interswitch trunks are four-wire circuits.

b. Another unique feature of AUTOVON is that all switching centers are toll-trunk like in design, and all are equal, i.e., there is no switching office hierarchy in AUTOVON such as is the case in the commercial network. This approach reduces the number of switches in a path and provides high quality four-wire circuits, at least between the end-to-end switched path, and in most calls, down to the local area DCO/PABX. Direct access subscribers are four-wire end-to-end, including the telephone instrument.

c. In addition to the advantages in circuit quality gained by using four-wire circuitry, the design parameters and interface requirements for connection to AUTOVON are closely controlled and are established to maintain the high quality requirements of the system. Tables 3-1 through 3-5 summarize these design parameters and interface requirements.

d. All access lines to an AUTOVON switch are, at least partially, part of the local area (DCO) subscriber loop plant and thus, an AUTOVON end-to-end connection consists of two such loops. Since poor loop design or maintenance can negate all of the quality built into the network, special attention is required; the overall criteria are summarized as follows:

(1)Non-Loaded Cable. The length of non-loaded cable used must not exceed 18,000 feet from the DCO/PBX to the user, regardless of gauge.

Table 3.1. Points of Interconnection

Connection	AUTOVON Switch	Main PBX 2/4 Wire	Tributary PBX 2/4 Wire	Station Equipment			
				4-Wire	2-Wire	4-Wire Data	Secure Voice
PBX access line	—	4-Wire —					
PBX tie trunk		—	2/4-Wire —				
User loop		—	4-Wire	—			
			—	4-Wire —			
		—	2/4-Wire	—	—		
			—	2/4-Wire	—		
Subscriber access line	—	4-Wire	—	—			
	—	—	4-Wire	—	—	—	
	—	—	4-Wire	—	—	—	—

Table 3.2. Loss Design Criteria

Type	Design Loss
Subscriber access lines	Uniform 6 dB
PBX access lines	0 dB
4-Wire main PBX to AUTOVON switch	0 dB
2-Wire main PBX to AUTOVON switch	4 dB
PBX tie trunks	
4-Wire main PBX to 4-Wire tributary PBX	0 dB
4-Wire main PBX to 2-Wire tributary PBX	4 dB
2-Wire main PBX to 2-Wire tributary PBX	3 dB
User loops	
To 2-Wire telephones	3 dB average to 7 dB maximum contingent on trunk arrangement, plus 4 dB where hybrid is used in user loop connecting to a 4-wire PBX
To 4-Wire telephones	Uniform 6 dB

Table 3.3. Equipment Application or Design Requirements

Echo suppressors	Located at switching centers on access lines to 2-wire PBX's and at 4-wire PBX's for tie trunks and users equipped with hybrids	
Telephone set		
Speech volume	Transmit:	
	Receive (See Note 1):	-11 VU, Sigma: 5 dB
		-19 VU (Maximum)
		-33 VI (Minimum)
Teletypewriter/data equipment levels (maximum in voice channel)	Data (Includes VFCT, AM, FSK, PSK)	- 9 dBm 0 (-8.7 dBm 0)
Hybrid		
Balance requirements (See Note 3). (When measured from hybrid to telephone set off-hook on PBX access line, PBX tie trunk, 2-wire user loop or special 2-wire secure-voice equipment.)	Echo return loss	8 dB (average) 6 dB (minimum)
	Singing point	5 dB (average) 2 dB (minimum)

Note 1: Does not include allowance for variations in talker volume and AUTOVON facility losses.

Note 2: For more than one sub-channel, this level shall be reduced by (10 log n) dB where n is the number of sub-channels.

Table 3.4. Loss on Overall Connections

Connection	Loss (dB)	
	Average	Maximum
4-Wire/4-Wire		
Voice	12 (See notes 1 & 2)	--
Data/secure voice	12 (See note 1)	--
4-Wire/2-Wire	13	17
2-Wire/2-Wire	14	22

Notes: 1. Uniform loss.
2. Does not include an additional 4 dB loss inserted in the receiving circuit of the 4-wire line adapter unit to compensate for removal of station hybrid loss.

Table 3.5. Balance Requirements

Measurement		Requirement			
From	To	Echo Return Loss (dB)		Singing Point (dB)	
Hybrid On	Terminated With	Average	Minimum	Average	Minimum
PBX access line PBX tie trunk 2-wire user loop Special 2-wire secure voice equipment	Station off-hook	8	6	5	2
PBX access line	PBX tie trunk (4-wire facilities) with 900 ohm + 2 MF termination at distant 2-wire PBX	22	16	15	11
PBX access line	PBX tie trunk (2-wire facilities) with 900 ohm + 2 MF termination at distant 2-wire PBX	18	13	10	6
PBX access line	PBX tie trunk hybrid with 4-wire legs terminated in 600 ohms	27	25	20	16
PBX tie trunk	PBX access line hybrid with 4-wire legs terminated in 600 ohms	27	25	20	16

(2)Loaded Cable. The type and length of loaded loops must be carefully controlled. The loop resistance must not exceed the supervision and signaling limits.

(3)Load-Coil Spacing. In order to achieve a smooth impedance characteristic, the loop must be fully loaded and the variations in load-coil spacing must not exceed +5 percent.

(4)Bridged Taps. Bridged taps or parallel operation of lines must be avoided. Bridged-taps create undesirable problems, such as impedance mismatch, consequent lower return loss and singing point, increased transmission loss, and the need for critical placement of taps in the case of loaded cable.

(5)Drop Wire. The length of drop wire should not exceed 500 feet from the cable terminal to the station equipment.

(6)Noise. The line noise must not exeed 20 dBa, as measured at the telephone, through the loop to the DCO/PBX.

(7)Noise Clicks. The magnitude of clicks must not exceed 53 dBa0, as measured by a Western Electric 6A Impulse Counter, or equivalent. Clicks are caused by inductive interference from switching, equipment, hits on line facilities, or switching operations while the connection is being established. High-level clicks of short duration are both a hazard and a source of annoyance to the subscriber or user. In areas with a high incidence of thunderstorms, special measures may be required to reduce the magnitude of the clicks.

(8)Room Noise. A room noise reference value of 50 dB Reference Acoustic Pressure (RAP) causes no impairment in a telephone connection. While it is not practical to design for variations in room noise, it is essential that special consideration be given to noisy locations.

(9)Amplitude Distortion. The insertion loss relative to 1000 Hz for voice and teletypewriter circuits (voice grade) must meet the objectives or minimums listed in Table 3-6.

(10)Transmission Variation. The allowable transmission variation at 1000 Hz is + 0.5 dB.

3-6. DISTORTION IN TELEPHONE TRANSMISSION CIRCUITS

The inherent imperfections in the various blocks comprising a transmission circuit alter the message signal as it travels from the source to the destination. The alteration of the message signal may be systematic or random (usually it is both). Examples of the former category are the frequency dependent attenuation and phase shift of the message through a transmission channel or an electrical device. Similarly, pulses may be always distorted in a certain way. Clearly, this type of distortion can be compensated for by using compensating networks. It is the random distortion incurred by the message that can cause annoyance for the telephone subscribers, and cause errors when data is transmitted over analog facilities. Examples of random distortion are white noise, impulse noise, cross talk, echoes, intermodulation products, drop-outs, etc. Being non-deterministic in nature, statistical and probabilistic models are used to understand the effect of random distortion. Steps are usually taken during the design and installation of

telephone circuits to reduce the effect of random distortion. Some of the common sources of random distortion are presented in the following paragraphs.

3-6.1. NOISE. Electrical transmission is accomplished by the motion of electrons in conducting materials and/or propagation of electromagnetic waves and thermal noise is associated with both. Thus, there cannot be electrical transmission without thermal noise. The effect of noise on the message signal depends upon the strength of the signal at various points in the telephone plant. In order to assure satisfactory transmission, precautions must be taken to assure an adequate signal-to-noise ratio at every point in the system. On almost all telephone connections, this ratio is maintained above 30 dB. It should be noted that if the signal strength becomes too low at some point in the system, the addition of more amplifier stages at the receiver will not improve the signal-to-noise ratio, because the noise will also be amplified with the signal. Thus, choice of repeater spacing is a crucial stage in the design of a telephone plant. White noise is not the only type of noise encountered in transmission systems. Single frequency noise, shot noise, and impulse noise also effect the message signal during transmission.

3-6.2. ECHOES. The term echo refers to the reflection of a wave from points in the transmission medium. This reflected energy, through amplification, may reach a certain point with sufficient magnitude and time delay such that it can be distinguished from the directly transmitted wave in voice transmission and can be quite annoying to the telephone users; in data transmission over telephone facilities, echoes are indistinguishable from

the directly transmited wave and can seriously impair the data transmission capabilities of a given channel by causing excessive error rates. The principal causes of echoes in wire line circuits are impedance mismatches in the medium, poor time equalization and impedance imbalance in hybrid coils. The amount of echo can be controlled by assuring an improved return loss at the terminal sets, and adding loss on the four- or the two-wire side. Echo suppressors are often used to reduce echoes. An echo suppressor is a voice operated electronic device used to block the passage of reflected signal energies. The blocking of reflected energy is carried out by inserting a high loss in the return four-wire path. CCITT recommends that echo suppressors be used if the mean round-trip propagation time exceeds 50 ms; ATT uses 4 ms as the threshold. It should be noted that an echo suppressor designed for voice must be disabled when the circuit is used for data transmission.

3-6.3. CROSS TALK. Cross talk refers to the disturbance created in one channel by the signals in other channels. It occurs between cable pairs carrying separate signals; in multiplexed channels, intermodulation products are a form of interchannel cross talk.

The nature of interfering cross talk is oten described as either intelligible or unintelligible. Cross talk between unlike channels is usually unintelligible because of all the changes the interfering signal undergoes in the different channel. The syllabic pattern of speech is usually retained even in unintelligible cross talk. In general, unintelligible cross talk is grouped with other noise-type interferences. It is the intelligible cross talk that is the peeve of many telephone users, and as such, cross talk objectives must be met during telephone plant design and installation.

There are three basic causes of cross talk in communication systems. First, the electrical coupling between transmission media can cause cross talk. Thus, it may originate in switching centers where large number of wires run parallel to each other around the exchange. It can be caused by capacitive as well as inductive coupling. Second, the nonlinear operation of repeaters in multiplexed channels can generate intermodulation products, which then appears as cross talk.

3-7. DATA TRANSMISSION ON VOICE CIRCUITS

At the present time, most data transmission is carried over the existing telephone facilities. Some networks for the exclusive use of data transmission have been constructed while others have been proposed. The DCS AUTODIN is an example of an existing exclusively data network; however, it currently utilizes telephone channels for all transmission paths. The usage of telephone lines for the transmission of data brought about an increased concern regarding the properties of these media and the effects of the media on the data signals. Imperfections on a voice channel may not seriously affect voice communications; however, these same imperfections may place serious limitations on data transmission. Table 3-6 summarizes the requirements for data service. Interface equipment is required to alter the digital signal in a way so as to make it compatible with the properties of the transmission media of the voice channel facility. Often, in addition to the interface device, delay and gain equalization are also required.

TABLE 3-6
Circuit Conditioning Criteria

Channel Conditioning	Attenuation Distortion (Frequency Response) Relative to 1004 Hz: Frequency Range (Hz)	Variation (dB) **	Envelope Delay Distortion: Frequency Range (Hz)	Variation (microseconds)
Basic (3002)	500-2500 300-3000	-2 to +8 -3 to +12	800-2600	1750
C1	1000-2400 300-2700	-1 to +3 -2 to +6	*1000-2400 800-2600	1000 1750
C2	*500-2800 *300-3000	-1 to +3 -2 to +6	*1000-2600 *600-2600	500 1500
C3 (access line)	*500-2800 *300-3000	-0.5 to +1.5 -0.8 to +3	*1000-2600 *600-2600 *500-2800	110 300 650
C3 (trunk)	*500-2800 *300-3000	-0.5 to +1 -0.8 to +2	*1000-2600 *600-2600 *500-2800	80 260 500
C4	*500-3000 *300-3200	-2 to +3 -2 to +6	*1000-2600 *800-2800 *600-3000 *500-3000	300 500 1500 3000
C5	*500-2800 *300-3000	-0.5 to +1.5 -1 to +3	*1000-2600 *600-2600 *500-2800	100 300 600
S3 (AUTOVON Access)	500-2800 300-3000	-0.75 to +1.5 -1.0 to +3.0	100-2600 600-2600 500-2800	80 250 500
4002 (Facsimile)	1200-2600	-3 to +3	1200-2600	600
4002 Cond. (Facsimile)	300-499 500-3000 300-3200	-2 to +6 -1 to +3 -2 to +6	1000-2600 800-2800	300 500

* These specifications are tariffed items.
** (+) means loss with respect to 1004 Hz.
(-) means gain with respect to 1004 Hz.

The provision of delay and gain equalization is called line conditioning when it is provided as part of the transmission circuit; many modern interface devices accomplish the required equalization automatically.

Figure 3.5 shows a schematic diagram illustrating conceptually the transmission of data in both directions on voice channel facility. For the sake of clarity, each functional block is shown separately; in actual systems, some of these blocks might be lumped together. The input/output device at each end may be a computer, a card reader, a teletype terminal, or some other data source/sink device. It also may perform such functions as coding and parallel/serial conversion and decoding and serial/parallel conversion.

The most commonly used digital data interface device for voice channel transmission is a modem. Modem is an acronym derived from modulator/demodulator, which are the two principal functions a modem performs. The function of the modem is to convert the digital information to analog signals which can be transmitted over the voice channel. At the receiving end, the modem converts the analog signal to the digital information which was originally transmitted. A variety of different digital modulation schemes are available. The most commonly used schemes are single sideband amplitude keying, frequency shift keying (FSK) and phase shift keying (PSK). Details of these modulation schemes are discussed later.

Devices called acoustic couplers are also used for transmitting data from low-speed devices The acoustic coupler is a type of modem that first converts the digital data signal into audible tones (usually one frequency tone for the one logic level and a second frequency tone for the zero logic level). In the opposite direction, it converts audible tones into digital data signals. The telephone handset fits into a special cradle in the

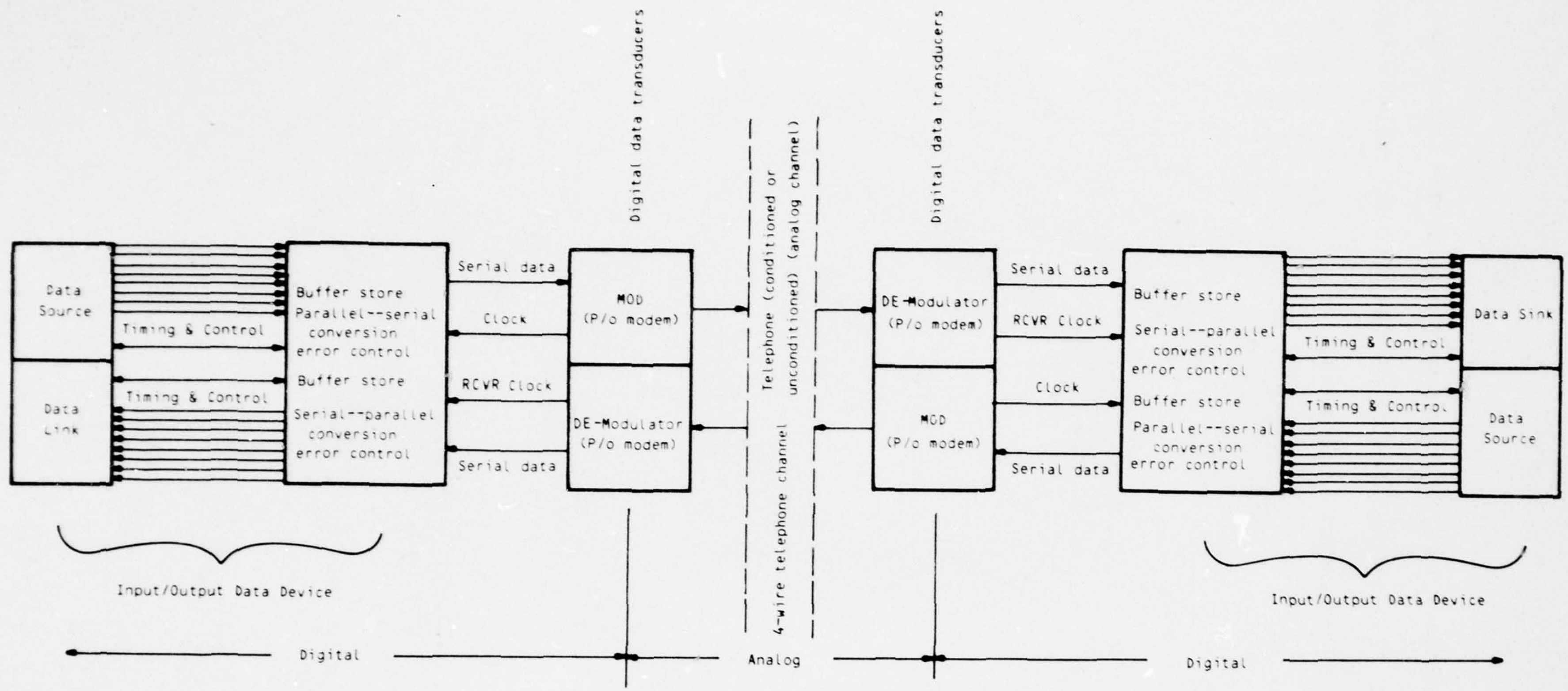


Figure 3.5. Functional diagram for digital data transfer on Analog System

acoustic coupling device and the tones act on the telephone transmitter in a manner similar to the voice. Analog devices such as facsimile (picture transmission) also utilize acoustic couplers.

SECTION IV

FUNDAMENTALS OF DIGITAL TRANSMISSION

4-1. GENERAL

With the increased use of automation and computers during the past two decades, the transmission of digital signals has become a major task in electrical communication. The ruggedness of digital transmission systems makes them attractive for use in the transmission of discrete messages - for example, data, written messages, symbols - as well as continuous messages, such as speech, video, etc. At the present time, most data is transmitted over the existing telephone network or other analog facilities that are based on the nominal 4kHz voice channel. Networks designed exclusively for data transmission have been constructed and others proposed. The DCS AUTODIN network is an example of an exclusive data transmission network; however, AUTODIN currently uses voice frequency channels for all transmission paths. AUTODIN II, an upgrade, includes wide bandwidth paths.

The steps involved in the transmission of discrete messages (or "data") are different from those of speech (or analog signals) over a voice channel facilities; and transmission of data signals over wire paths in a digital mode is different from transmission over base wire in a 4 KHz voice channel mode. This section first contrasts digital with analog transmission; this is followed by a discussion of the fundamentals of data transmission.

4-2. ANALOG vs. DIGITAL TRANSMISSION.

In analog transmission systems, the signal applied to the transmission medium varies continuously with message waveform. The signal amplitude, frequency or phase, varies continuously with time. In digital transmission, on the other hand, the signal is discrete in state and time. The simplest

example of a digital signal is a pulse train, where either a pulse or a space (no pulse) is transmitted. Such signals, where the signal states are limited to two values - 0 or 1 - are referred to as binary signals. Binary signals are easy to generate and process, and are, therefore, most commonly encountered. Other signals waveforms with multiple, but discrete, signal states are also used in digital transmission and are referred to as m-ary signals.

An important parameter in analog transmission is the signal or system bandwidth. Similarly, in digital transmission, the signals are characterized in terms of Baud, which measures the number of shortest equal-length symbols per second capable of being transferred by the channel. In binary transmission, bauds and bits-per-second can be synonymous, if all pulse lengths are equal. Bit rates for commonly encountered systems may range from 40 bits/sec to several megabits/sec. For example, manual keyboard devices generally operate from 45 to 300 bits/sec, while automatic transmission devices may involve bit rates of up to 50 Kb/s.

The transmission reliability in analog voice systems is measured in terms of signal-to-noise ratio and intelligibility at the receiver. In digital transmission, the criterion for measuring reliability is the bit error probability versus S/N at the receiver. Just as the intelligibility and acceptable signal-to-noise ratio in analog systems varies from system to system, depending on the particular system application, similarly the acceptable bit error rate versus S/N depends on what the system is used for. Bit error probabilities may range from 10^{-3} (one error bit in 1000 transmitted bits) to 10^{-12} (one error bit in 10^{12} transmitted bits).

In summary, signaling speed in terms of bits-per-second and Baud, and reliability of transmission in terms of bit error probability versus S/N are

the basic parameters of digital transmission systems, analogous to bandwidth, intelligibility and signal-to-noise ratio in analog voice transmission systems. The paramount concern in analog transmission systems is how well the received waveform resembles the transmitted waveform. Thus, the shape, and concomitantly the frequency content, of the received signal is of consequence. In digital transmission, the main concern is the accuracy with which the signal, containing the message in coded form, is delivered. Here, the received sequence of pulse states is of consequence - the shape of an individual pulse is not important as long as its signal state is within the range of values which permits a correct decision by the receiver.

Instead of amplifying the transmitted signal at regular intervals - as is done in analog transmission - the repeaters in a digital transmission systems transmit a new waveform almost "identical" to the input waveform. This process is called <u>regeneration</u>. The ability to regenerate the transmitted signal at regular intervals without any significant errors results in a better utilization of the noisy media, and is a most important advantage of digital transmission.

4-3. BASEBAND DATA TRANSMISSION.

The digital signal, representing a written message or data, consists of a string of discrete rectangular pulses. Some of the commonly used digital signal forms are described below. Figure 4-1 shows the common formats.

4-3.1. UNIPOLAR (NEUTRAL). This is an ON-OFF sequence; it is called unipolar because there is only one polarity of voltage or current. One signal level is zero, the other level is the maximum signal level. In m-ary neutral signalling, one level is the zero signal level, another is the maximum

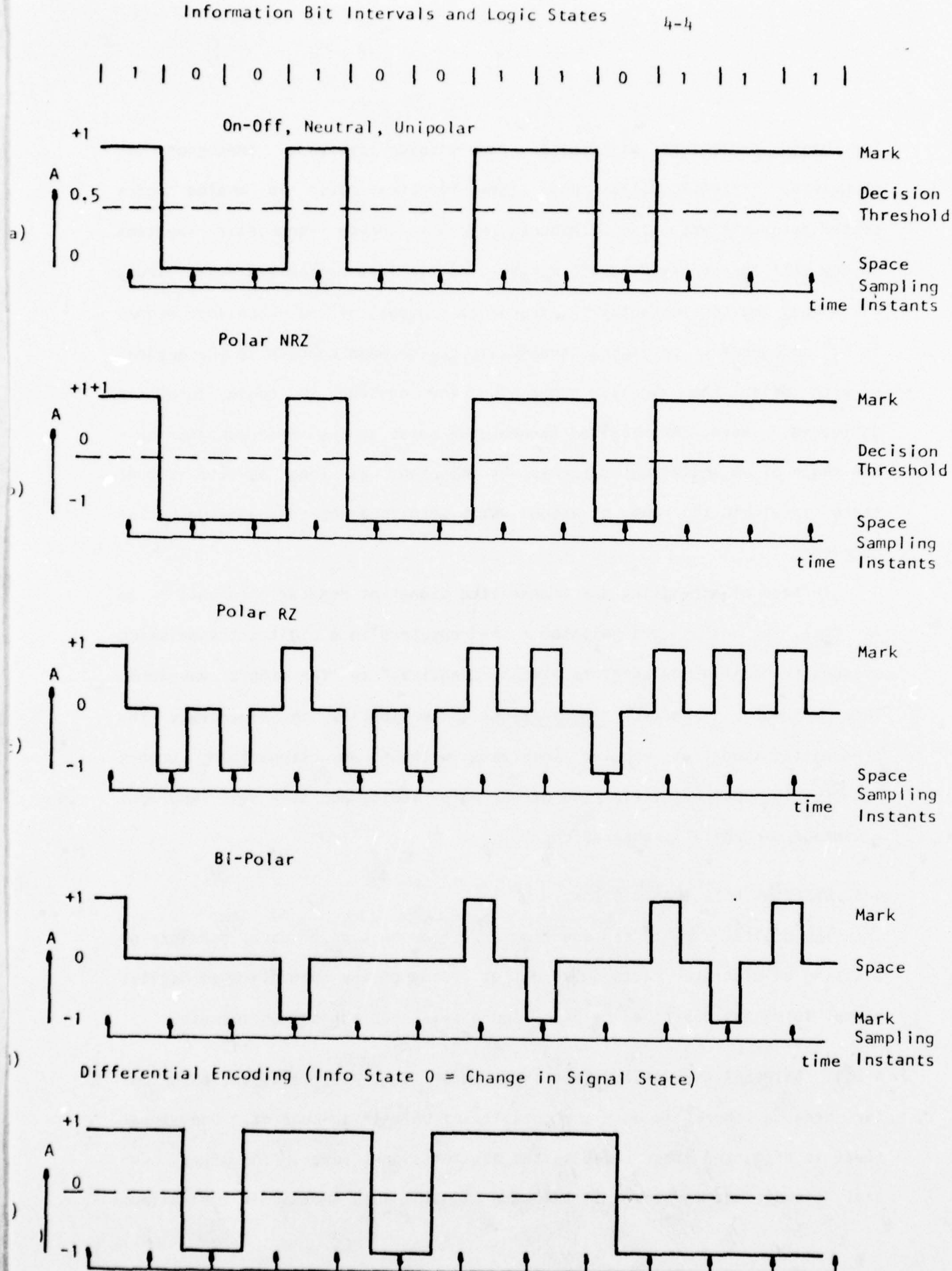


Figure 4.1. Signal Formats

signal level, and all other signal levels are between the zero and maximum level.

4-3.2. POLAR NON-RETURN-TO-ZERO (NRZ) signals are those in which one state is represented by plus maximum signal level and the opposite state by a numerically equal negative maximum signal level. In m-ary signaling the remaining states are symmetrical plus and minus signal levels between zero level and the maximum levels, and in the case of an odd number of levels, the zero level is utilized.

4-3.3. POLAR RETURN-TO-ZERO (RZ) SIGNALS are those in which the information bit rate is doubled or conversely the pulse interval is halved. Each information bit is presented to the line as two signal levels. One state is represented by the maximum positive level for the first half of the information bit interval and as zero signal level for the second half of the information bit interval. The opposite state is represented in the first and second halves of the information bit interval by the maximum negative and zero signal levels, respectively.

4-3.4. BI-POLAR SIGNALS are those in which one state is always represented by the zero signal level and the opposite state is represented by either the maximum positive or maximum negative signal level for the first half of the information bit interval and the zero level for the second half. As this state appears in the information bit stream, it is assigned the positive and negative values on an alternating basis.

4-3.5. DIFFERENTIAL SIGNALING is a scheme in which one information state is

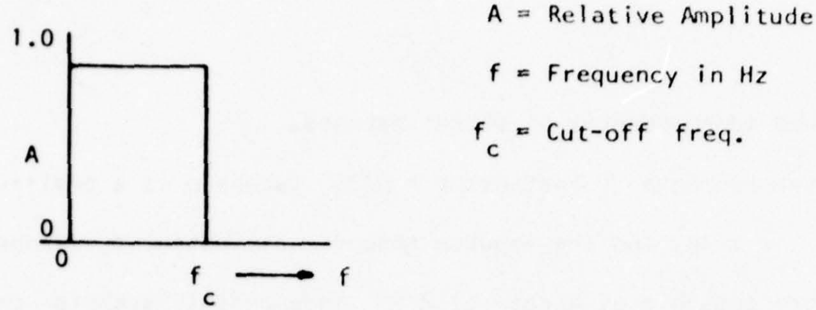
represented by a change in signal level from the maximum level in one direction to the maximum level in the opposite direction, halfway between adjacent sampling instants, and the opposite information state is represented by no change in signal level. The resulting pulse train resembles a binary polar NRZ signal and m-ary encoding can be accomplished after the differential coding operation in the same manner as with the polar NRZ signal.

4-3.6. DEFINITIONS: In order to evaluate the relative advantages of the various signalling formats, one must have a standard basis for a comparison. However, in a review of current literature on the general subject of data transmission, one is soon aware that comparison of the data presented is not possible without first performing some sort of normalization of that data. This confusion comes about for the most part by the use of Nyquist and Shannon terms intermixed with adjusted terms without specifying which is which and when. In order, hopefully, to avoid this same problem in this report, this section sets down and defines the terms to be used in comparing the utility of the various signalling techniques.

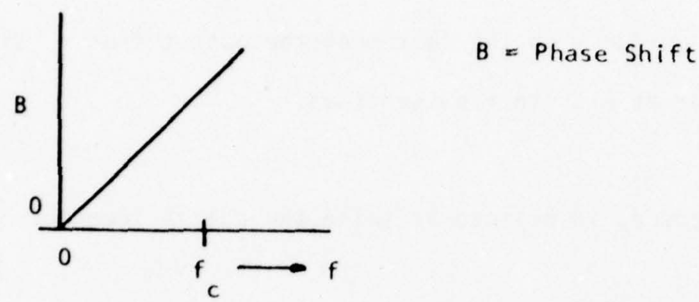
a. Nyquist I Signalling. Unfortunately, there is no common agreement for the Nyquist bandwidth or channel capacity other than for signaling with rectangular pulses applied to an ideal low pass filter. These agreements are based on Nyquist's first criterion and thus might be called the Nyquist I bandwidth, rate and interval (Figure 4.2).

The first criterion requires that the signal have equally spaced axis crossings in the impulse response.

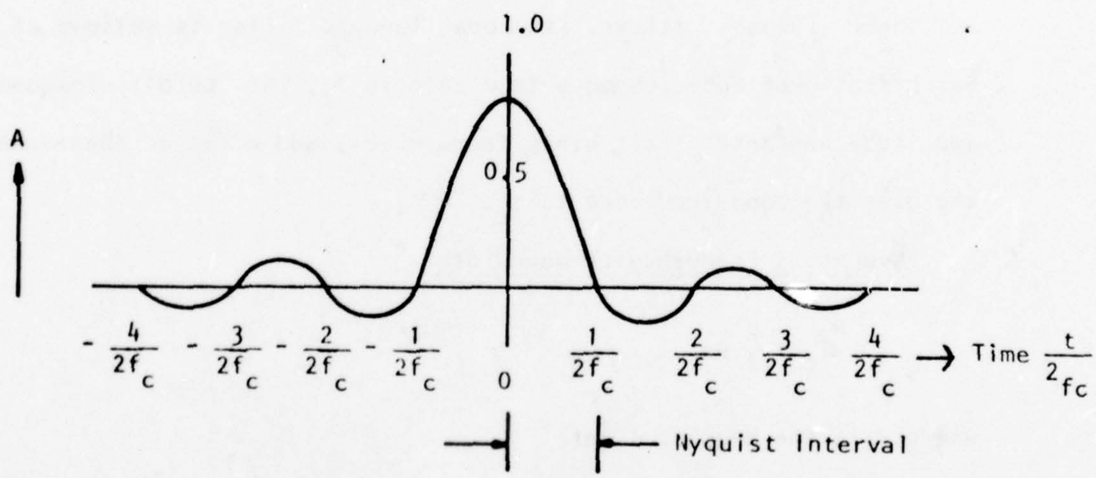
Nyquist has shown that if rectangular impulses are applied to an ideal low-pass filter at instants separated by $1/2\ f_c$ seconds (where f_c is the cut-off frequency of an ideal lowpass filter), the response to these im-



Amplitude Function of an Ideal Lowpass Filter



Phase Function of an Ideal Lowpass Filter



Impulse Response of Lowpass Filter

Figure 4.2. Nyquist I Signalling

pulses can be observed independently of those instants.

Since the impulse response at instants $t = n/2f_c$ (where n is a positive or negative integer) is zero, and the impulse response at instant t_o is one, signaling is therefore possible at a rate of $2 f_c$ independent symbols per second.

This property depends uniquely on the fact that the output from a single pulse has zero amplitude at all other pulse times.

Thus, the Nyquist I,

Rate in symbols per second, is defined as twice the cutoff frequency of an ideal lowpass filter.

Interval, in seconds, is defined as the reciprocal of twice the cutoff frequency of an ideal lowpass filter.

Bandwidth or Frequency, in Hertz, is defined as the cutoff frequency of an ideal lowpass filter. An ideal lowpass filter is defined as one which has a flat amplitude response from zero to f_c, the cutoff frequency, zero amplitude response at all other frequencies, and a linear phase characteristic over the band from zero to f_c.

Nyquist I Frequency or Bandwidth

$$f_c = \frac{R}{2} \text{ Hz} \tag{4-1}$$

where R is the Nyquist I rate

Nyqist I Interval

$$t_o = \frac{1}{2f_c} \text{ seconds} \qquad (4\text{-}2)$$

Nyquist I Rate

$$R = 2f_c \text{ Baud} \qquad (4\text{-}3)$$

$$R = 1/t_o \text{ Baud} \qquad (4\text{-}4)$$

where Baud is an expression for equal length symbols per second.

Extension of Nyquist I Definitions to m-ary Signaling:

A common misconception on the part of many communicators is that the maximum or Nyquist signaling rate is equal to two bits of information per Hertz of available bandwidth. But, Nyquist's first criterion states that at most, we can transmit two independent pulses per Hertz of bandwidth. These pulses can be binary (bits), but are not necessarily so. Nyquist's first criterion makes no such restriction and thus we can also transmit m-ary independent pulses at the rate of two pulses per second per Hertz of bandwidth.

If we transmit two independent pulses through an ideal lowpass filter, one with amplitude a_o at time t_o, the other with amplitude a_1 at time $t = 1/2f_c$, the response at time t_o will be a_o and at time $t = 1/2f_c$ will be a_1.

For a binary signal the Nyquist rate

$R = 2\ f_c$ or $R = 1/t_o$ baud = BPS

A general expression for the Nyquist channel capacity in bits per second which includes all classes of symbol encoding, is given by:

$$C = 2f_c n \text{ or } C = (1/t_o)n \text{ BPS} \tag{4-5}$$

where n is the effective multiplier of the signaling rate in a given bandwidth.

For conventional m-ary encoding, the relationship of n is :

$$m = 2^n \tag{4-6}$$

where m is the number of available independent symbols in the encoding scheme.

Thus, for conventional m-ary encoding the Nyquist capacity is given by:

$$C = 2f_c \log_2 m \text{ or } C = (1/t_o)\log_2 m \text{ BPS}$$

The maximum signaling or symbol rate (as opposed to the information rate in BPS) is given by:

$R = 2\ f_c$ baud or $R = (1/t_o)$ Baud

For binary signaling R (Baud) = C (BPS)

For general m-ary signaling;

$$R = \frac{2f_c}{n} \text{ Baud} \tag{4-7}$$

For conventional 2^n m-ary signaling;

$$R = \frac{2f_c}{\log_2 m} \text{ Baud} \tag{4-8}$$

The bandwidth efficiency or figure of merit is the number of information bits transmitted per Hertz of bandwidth and is given by:

$$F = \left[\frac{R}{f_c}\right] n \text{ Bits/Hz for the general case and} \tag{4-9}$$

$$F = \frac{R}{f_c} \log_2 m \text{ Bits/Hz for the conventional } 2^n \text{ m-ary case.} \tag{4-10}$$

It should be noted that all of the foregoing equations are directly related to Nyquist's first criterion where the bandwidth f_c is the cutoff frequency of an ideal lowpass filter which is equal to one-half the symbol rate in bauds.

Although the ideal lowpass filter is not a realizable filter, there is no common agreement for the Nyquist rate or frequency for other than an ideal lowpass filter, and thus it has become conventional to use it as a reference when comparing digital signaling schemes on a theoretical basis.

b. Nyquist II Signaling. Even if the ideal lowpass filter were realizable, a slight error in cutoff frequency, pulse rate or receiver sampling instant, would result in an overlap of the pulse tails, which represent a divergent series and add up to large values, causing interference between symbols.

Nyquist's second criterion requires equal times between transition values to remove intersymbol interference at the instants half-way between adjacent samples.

If the response to a regularly spaced pulse train at the Nyquist I rate is measured at instants halfway between adjacent impulses, the response will be f_c times the sum of the two adjacent impulse weights. We obtain values proportional to the average signal samples.

In the case of binary signaling the average of adjacent values is either zero, one-half, or one, depending on whether the pair is two zero's, a

one and a zero, or two one's.

Nyquist has shown that one can simultaneously satisfy both the first and second criterion, preserving both nulls and transitions, by substituting a raised cosine filter, with even symmetry about the Nyquist I frequency, for the ideal lowpass filter (Figure 4.3).

Another advantage of the raised cosine filter is that transient tails die away rapidly and pulse lengths different from the exact Nyquist interval (as defined by the frequency about which the raised-cosine shaping is symmetrical) can be transmitted with small amounts of intersymbol interference.

The cost of this freedom from intersymbol interference is a bandwidth twice that of the ideal lowpass filter. On the basis of freedom from intersymbol interference, the Nyquist II rate or channel capacity is the same as the Nyquist I rate when f_c is defined as the frequency about which the filter has a vestigally symmetric cutoff characteristic.

Thus, the <u>Nyquist II</u>

Rate $R = 2f_c$ Baud (4-11)

<u>Interval</u> $t_o = 1/2f_c$ (4-12)

<u>Bandwidth or Frequency</u>

$$f_{cII} = f_c + f_x \tag{4-13}$$

where $f_x = f_c$ in the case of the raised-cosine filter and thus

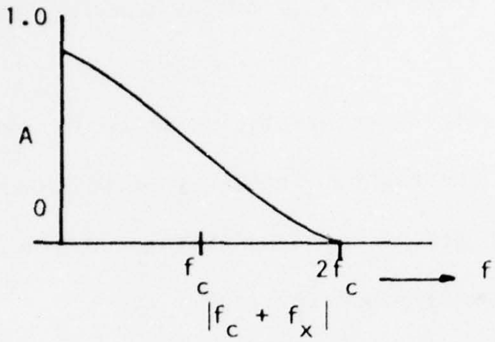
$$f_{cII} = 2f_c$$

often referred to as the bit rate bandwidth since the Nyquist II rate

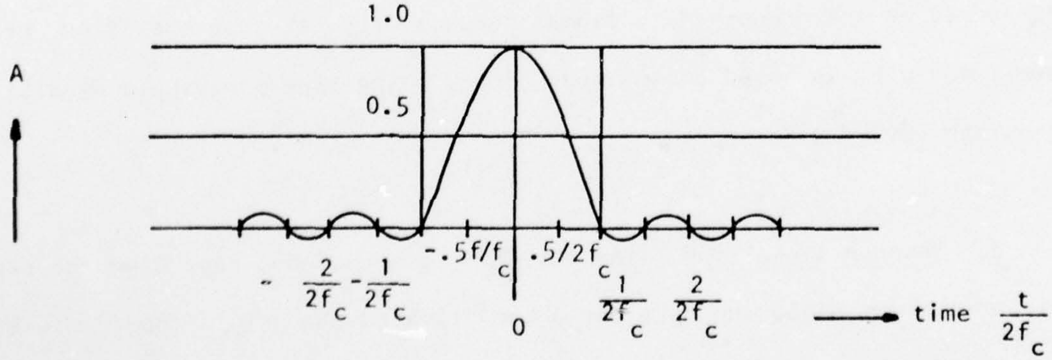
A = Relative Amplitude

f = Frequency in Hz

f_c = Frequency about which symmetry exists



Amplitude Function of a Raised Cosine Filter



Impulse Response of Raised Cosine Filter

Figure 4.3. Nyquist II Signaling

$$R = f_{cII} \text{ Baud} \tag{4-14}$$

The bandwidth ($f_c + f_x$) of a practical channel filter network, in all cases, is greater than the Nyquist I frequency (f_c) and is usually less than the Nyquist II frequency ($2f_c$).

Channels with a gradual cutoff characteristic could be regarded as having a Nyquist II bandwidth up to the highest frequency at which there is any significant response; or we could define $2f_c$ baud as the xdB Nyquist I rate (i.e., a loss of xdB at all frequencies greater than f_c).

Some definition is absolutely necessary for comparison of signal performance in relation to the Nyquist rate.

An alternate definition of the Nyquist rate could be the maximum signal rate possible over a channel with some x% of intersymbol interference.

We avoid intersymbol interference if we signal at the rate of $2f_c$ or f_{cII}, but there is no real contradiction if by accepting some intersymbol interference we signal successfully at $x(f_c)$ where $x > 2$ or $x(f_{cII})$ where $x \leq 2$.

In this report, for the purpose of comparing signal processing schemes, the required bandwidth (or conversely the bandwidth efficiency) will be considered to be that required to signal successfully (at some specified error rate) and will be based on an input binary pulse stream having a Nyquist II bandwidth (BPS = Hz).

c. Shannon Signalling. The foregoing discussions, regarding the capacity or maximum signaling rate for a band limited channel, without intersymbol interference, are based on a noiseless channel.

For noisy channels we must refer to Shannon's theorem to obtain the capacity of the same ideal lowpass filter channel.

Shannon's ideal system capacity is given by:

$$C_N = f_c \log_2 (1+\frac{S}{N}) \text{ BPS} \tag{4-15}$$

where C is the maximum number of bits per second which can be received with arbitrary small error over the band limited channel with fixed average power and additive white Gaussian noise (Fig. 4.4),

f_c = the Nyquist I bandwidth

S = average signal power;

N = average noise power in the Nyquist I bandwidth;

the spectral density of the noise is constant and equals $\frac{N}{f_c}$ and the probability distribution function of the noise is Gaussian. Also,

$$\frac{C_N}{f_c} = \log_2 (1 + \frac{S}{N}) \text{ bits/Hz} \tag{4-16}$$

There are specific conditions to the Shannon theorem which must be considered if it is to be utilized in any comparison of various signaling schemes.

(1) There are no restrictions on the complexity of the transmitting or receiving equipment, or associated coding schemes.

(2) The bits transmitted per Hertz of bandwidth include all bits; i.e., synchronization bits, framing bits, parity bits, etc.

(3) The bandwidth is the Nyquist I frequency f_c, the cutoff frequency of an ideal lowpass filter.

(4) The noise is white noise; i.e., it has a flat power density spectrum. It should have units of watts per unit bandwidth, but it is common practice in noise theory to consider the (amplitude)^2 as the unit of power. This gives a power density spectrum in units of (amplitude)^2/Hz. This in-

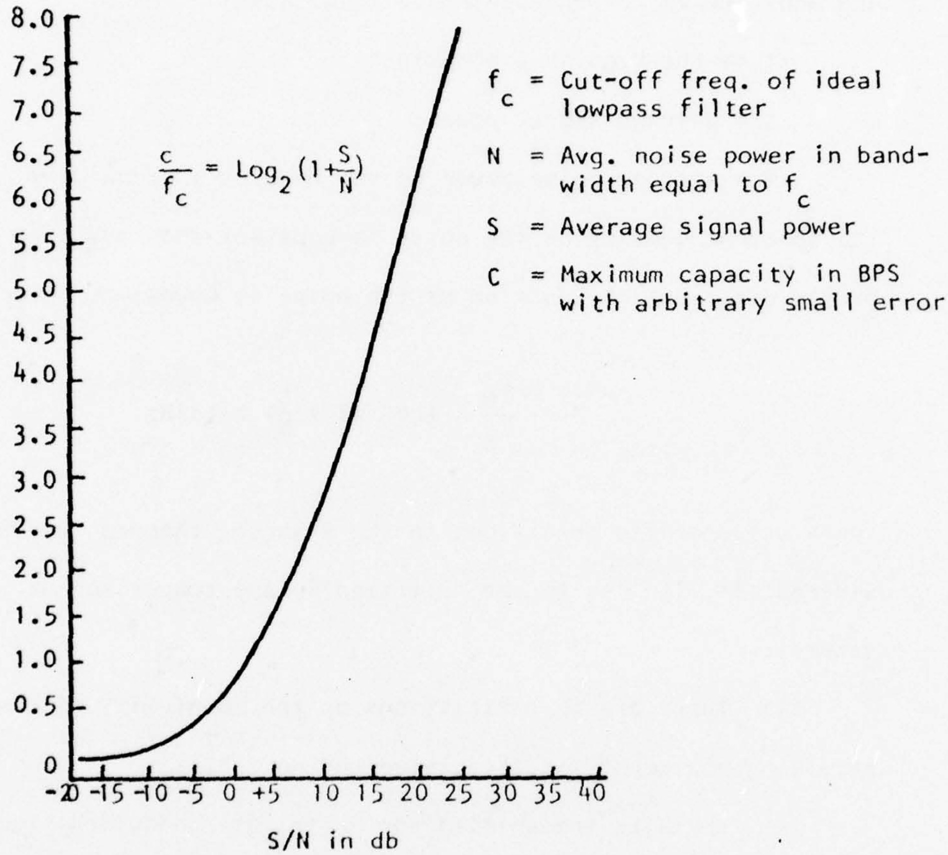


Figure 4.4. Shannon Channel Capacity

consistency is unacceptable to some engineers and they reconcile the difference by assuming a one ohm load resistance. But, this is really unnecessary if one considers that the noise power available from a resistor is given by:

$$p = \frac{en^2}{R} = \frac{4ktBR}{4R} = KTB \qquad (4\text{-}17)$$

where e_n is the open circuit voltage,

R is the internal resistance which in itself is noise free,

k is Boltzman's constant of 1.38×10^{-23} joules per degree Kelvin,

T is the absolute temperature $273.16^{\circ}K$ plus degrees centrigrade above zero, and

B is the bandwidth over which e_n is measured.

The noise power KTB is independent of R and in the case of white noise, the noise power density KT is also independent of R.

(5) The noise has a probability density function which is Gaussian and which is characteristic of many naturally occurring random disturbances. It is important not to confuse the Gaussian probability density function with the output of a Gaussian filter. The filter has an impulse and frequency response shaped like a Gaussian curve. The output of such a filter may indeed have a Gaussian probability density function also, but an arbitrary signal having a Gaussian probability density function may have a power density spectrum which bears no resemblance to the frequency response of a Gaussian filter. This is, of course, the case with the Shannon specified noise. It is also important to recognize that Gaussian noise does not have to be white noise.

(6) If the noise is either non-white or non-Gaussian or both, the value of C_N would be greater than that given by the Shannon theorem.

d. Noise Power

The value of the noise power in a system is the average noise power at the output of the receive filter or input to the detector. However, it is more convenient for relative comparison purposes, to use the average noise power in a specified bandwidth. In the case of white Gaussian noise, by definition, the mean total noise power is directly proportional to the bandwidth.

$$\text{Mean total noise power} = KTB = N_o B, \text{ where}$$

$$N_o = KT \text{ the noise power density } V^2/Hz \qquad (4\text{-}18)$$

$$B = \text{bandwidth}$$

One convenient reference is the average noise power in a bandwidth equal to the symbol rate (Nyquist II). Another is the average noise power in the Nyquist I bandwidth corresponding to one-half the symbol rate.

For these references, the noise in the Nyquist I bandwidth is 3 dB less than in the Nyquist II bandwidth.

A third reference used by many authors is the average power is a bandwidth equal to the bit rate. In the binary signaling case, this is the same as the symbol rate for the Nyquist II bandwidth, but in the m-ary case the relationship is a function of the encoding scheme.

In this report the noise power bandwidth will be the same as the Nyquist II signaling bandwidth for baseband signaling. For other than baseband signaling (Section V) it will be the bandwidth required to support a Nyquist II baseband.

e. Signal Power

A variety of references can also be used for signal power. One reference is the average signal power for a random pulse train. Another is the

average signal power for some specific pulse train. A third is the peak value of signal power since the ratio of peak-to-average signal power varies with the modulation method and thus becomes the real constraint in many practical systems. A fourth reference which is similar in nature to the noise power density V^2/Hz is the average power per pulse; or the average power per bit.

Assuming a random pulse train, in the binary case, the average total signal power is simply the average power per pulse, or per bit, times the pulse or bit rate.

In the m-ary case the power per bit relationship to the power per pulse or total power is a function of the encoding scheme.

In the design of digital transmission systems the maximum available peak power, bandwidth and information rate are normally fixed constraints, whereas the engineer has relative freedom in the selection of the encoding scheme (m-ary encoding). It therefore appears that the value of the energy per information bit is useful in that it also provides a description of the encoding scheme when it is compared to an equal power, equal bandwidth binary or uncoded scheme.

f. S/N Ratio

Based on the foregoing discussion we will make comparisons on both an average total signal power and maximum steady-state signal power to noise ratio basis, and on both an equal pulse rate and equal bit rate basis.

g. Basis for Comparison of the Various Signaling Schemes

In order to present a useful set of performance comparison data, it is reasonable to compare each scheme to each other scheme, or to some reference scheme, on the basis of error performance as a function of each of the sys-

tem parameters.

Error performance can be stated as the "probability of error" or as the "Bit Error Rate." The probability of error is simply the probability that a bit will be received in error for each bit transmitted.

The bit error rate is a statement that x bits will be received in error for y bits transmitted, and having a discrete limit (y), is more convenient for use in actual bit error performance measurements.

If the number of bits transmitted is sufficiently large, the bit error rate (BER) approaches the probability of error, and on this basis one finds the two terms used interchangeably in the digital community.

The probability of error depends on three considerations; the transmission media, the modulation scheme, and the demodulation process.

In the process of selecting a signaling scheme for use in a specific environment, one would normally compare the various available schemes on the basis of ideal performance within the general signaling environment and then subject those which show the most promise to further analysis within the constraints of the specific real environment, and the less than ideal performance of the transmitter and receiver.

The general signaling environment is specified in terms of bandwidth, level of additive white Gaussian noise, and peak or average total signal power.

The specific real environment would be further specified in terms of channel non-linearities, signal fading, impulse noise, and interference from other sources.

We shall make the following assumptions in regard to the ideal system for each signaling scheme:

(1) The channel is bandlimited.

(2) The composite function of the transmitter filter, channel, and receiving filter provide a raised-cosine pulse shape output to the detector.

(3) There is no intersymbol interference.

(4) The channel is a non-fading channel.

(5) The additive noise is white Gaussian noise.

(6) The transmitter and receiver are optimum.

(7) The digital signal is random and all signal values have equal probability and are statistically independent.

(8) In double sideband systems the spectral density function is one in which the power is divided equally between plus and minus frequencies.

Under this set of assumptions the probability of a pulse error is a function of the ratio of the received pulse amplitude to the rms noise value at the sampling instant and has been shown to be:

$$P(e) = \frac{1}{2}(1-\mathrm{erf}\sqrt{m}) \text{ (for an ideal receiver)} \qquad (4\text{-}19)$$

where erf is the error function, and m is a statement of the signal-to-noise ratio which must be specified in terms of bandwidth and bit rate.

In m-ary modulation schemes the relationship between the probability of a pulse error and an information bit error is not straight forward. Although one can specify m in terms of the relationship of the number of bits encoded per pulse, the received pulse error can have any one of the

other (m-1) values; with high signal-to-noise ratios, only the adjacent levels are the most likely to be received in error, and with low signal-to-noise ratios, all levels are likely. In addition, the encoding scheme determines the value of each pulse.

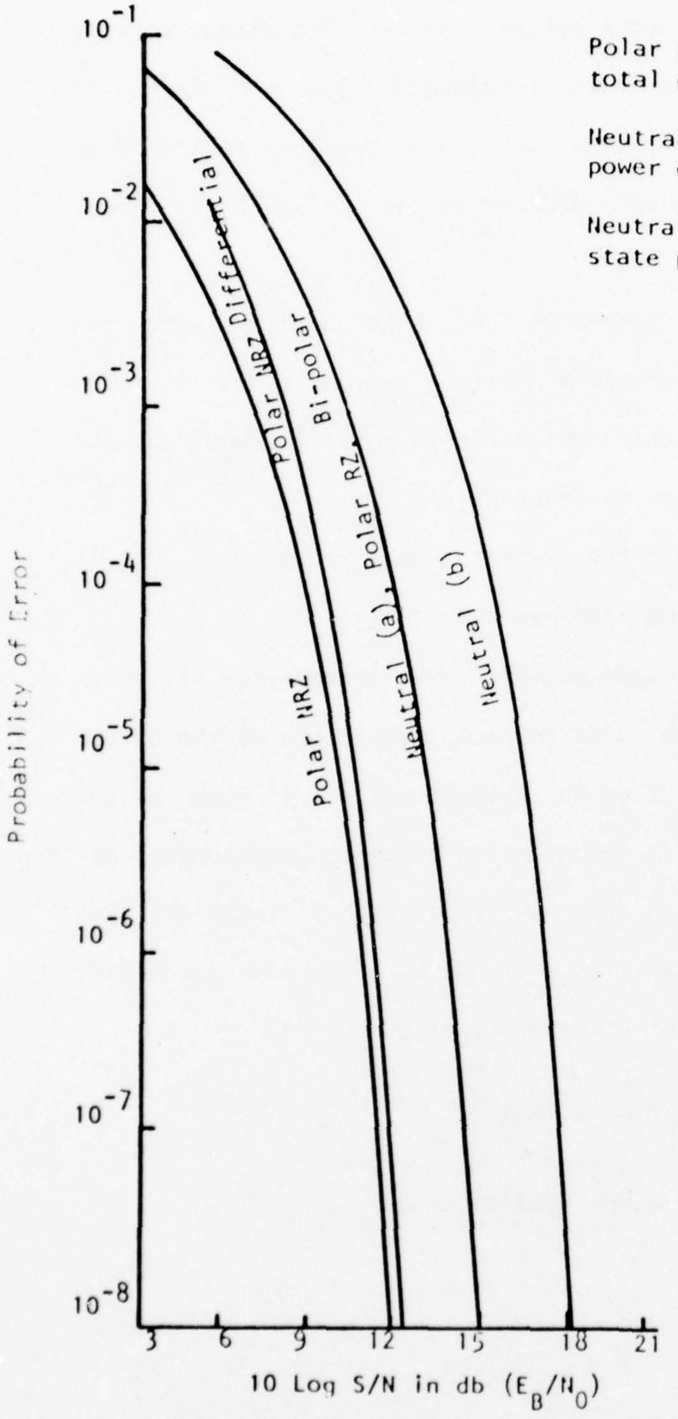
With straight forward m-ary encoding the bit error probability is greater than the pulse error probability since mutilation of a single pulse can result in more than a single bit error. However, with Gray coding, there is only one digit difference in adjacent levels (00, 01, 11, 10, 110, 111, 101, 100) and thus for high signal-to-noise ratios, the probability of a bit error closely approaches the probability of a pulse error.

For this reason, which offers some simplicity in system performance comparison, we will assume Gray coding for all m-ary schemes, unless otherwise stated in the specific analysis of a given system, and assume a single bit error per pulse decision error.

4-3.7. IDEAL PERFORMANCE OF BASEBAND SIGNAL FORMATS (Fig. 4.5)

(a) Unipolar and Polar NRZ: In terms of required bandwidth, ideal unipolar and polar NRZ signaling are equal. Both can be specified as $f_{cII} = 2f_c = R$ for Nyquist II signaling. In terms of bandwidth efficiency or figure of merit, both can be specified as $F = 1$ bits/Hz for Nyquist II signaling. White Gaussian noise power is the same for each of the signal formats. In the Nyquist II case, the average total noise power is twice that of the Nyquist I case.

However, in terms of signal power the two formats are not equal when considered in terms of equal peak-to-peak signal amplitude A.



Polar NRZ curve is the same under average total or maximum steady-state power limit.

Neutral curve (a) is under an average total power constraint.

Neutral curve (b) is under a maximum steady-state power limit.

Assumes raised cosine pulse spectrum additive white Gaussian noise, and Nyquist II signalling

Figure 4.5. Probability of Error for Baseband Signals (Binary with equal pulse-rate)

It has been shown that the spectrum of the special case of a binary train of alternate on-off squarewave pulses consists of a direct current component plus components consisting of the fundamental and odd harmonics thereof, in which one-half of the average total signal power is contained in the DC component. The equivalent polar spectrum is the same with the direct current component omitted.

If uncurbed binary rectangular pulses with no band limiting are assumed, the peak signal power per ohm is A^2 for the neutral signal and $A^2/4$ for the polar NRZ signal, showing that 6 dB more peak power is required for the neutral signal to obtain the same decision margin.

On an average signal power basis the neutral signal requires $A^2/2$ and the polar signal $A^2/4$, showing a 3 dB difference.

In the case of additive white Gaussian noise, the probability of error is related to the probability of the noise component amplitude at the output of the receiving filter exceeding A/2 at the pulse centers. If the binary wave is positive, with respect to the decision threshold, the noise component must be negative to produce an error and vice versa. Thus, at any one sampling instant we are concerned with only one polarity of noise peaks.

Using the general expression for probability of error:

$$P(e) = \frac{1}{2}(1-\operatorname{erf}\sqrt{m})$$

we can define m as follows for the neutral and polar NRZ signals:

<u>Polar</u> <u>(Binary)</u>:

$$m = S_p/N \qquad (4\text{-}20)$$

where S_p is the average total signal power and N is the average total noise power in a Nyquist II (pulse rate) bandwidth; thus

$$m = S_p/2N_of_c \tag{4-21}$$

<u>Neutral</u> <u>(Binary)</u>

$$m = S_n/N \tag{4-22}$$

where S_n is the average total signal power, which is twice that of the polar signal S_p, and N is the average total noise power in the pulse rate bandwidth.

$$m = \frac{S_p}{N_of_c} \text{ for equal probability of error} \tag{4-23}$$

The E_B/N_o ratio is equal to the S/N ratio for Nyquist II signaling (since the bit rate equals the bandwidth) for equal probability of error in the binary neutral and polar NRZ cases.

These signals utilize the pulse interval efficiently in that the entire interval contains signal information. Timing information is contained in the signal transitions or decision level crossings and thus is a function of the transition density which in turn is a function of the pulse train pattern.

Neutral signals, in addition to the 3 dB S/N disadvantage, have an additional disadvantage when compared to polar NRZ signals. In a neutral system with an A/2 slicing level, a 10% decrease in A causes an approximate 7% pulse interval error (spacing bias) since the slicing level is no longer at A/2. In a polar NRZ system the slicing level remains at zero and the mark-space interval symmetry is maintained as long as the mark-space amplitude symmetry is maintained.

For these reasons, neutral signaling is seldom used except for very short cable paths such as between devices in the same general physical area.

b. Polar RZ: Polar RZ signals have a broader frequency spectrum for a given information signaling rate because of the fact that each information pulse is divided into two shorter pulses. As a result of the wide spectrum, less of the energy is at zero and the very low frequencies which makes it feasible to transmit these signals through transformers, which are the primary coupling devices in the vast cable and microwave frequency division multiplex networks. These signals essentially waste one-half of the bandwidth in that only one-half of the bit period contains signal information. However, if the zero amplitude intervals are considered to be a third level and appropriate slicing levels are implemented, a zero crossing is assured during each information bit period and timing information can be readily extracted and the transition density is not dependent upon the information pattern.

c. Bi-Polar: Bi-polar signals also have a broader frequency spectrum and have essentially zero energy at DC and the very low frequencies. Most of the energy is concentrated at a frequency equal to one-half of the information bit rate. These signals also essentially waste half of the bandwidth and if the zero crossings are utilized to extract timing information, it is noted that the transition density is a function of the number of information one's transmitted. The advantages of this signal format are the relatively high concentration of energy in the center of the band which permits transformer coupling and less than ideal high frequency cutoff in the channel plus the format requirement of alternating polarity of one's permits in-service error detection.

The 3 dB los in margin against noise suffered by polar RZ and bi-polar signals is not significant when these formats are utilized over cable facilities which have regenerative repeaters at relatively close intervals.

d. Differential Encoding: Differential encoding, as the term implies, is not a form of pulse format. The differentially encoded pulse train may be transmitted in any of the other pulse formats. It is included here in order to determine its effect on error probability.

One is tempted to intuitively suggest that since the information is represented by the direction and distance in the change in amplitude between two sampling instants (in two information bits) that an error in decision at one instant (bit) would also produce an error at the following instant (bit). However, it has been shown, analytically and experimentally that this is not necessarily so. When one considers that the probability of a decision error, due to the noise, at the second sampling instant is the same as the probability at the first instant, and that such a second error would complement the first error, it seems logical that the degradation due to differential encoding would be some quantity less than two and more than one times the error probability of an uncoded pulse train.

For simplicity, however, this report will utilize the conventional concept that differential encoding does produce bit error pairs for each error pulse, unless specifically stated otherwise.

e. m-ary Signals (Fig. 4.6). To make decisions between more than two levels requires additional decision thresholds on the basis of

$$L = (m-1) \tag{4-24}$$

where L is the number of decision thresholds and m is the number of levels.

In the quaternary case (four levels), three thresholds are required, so that for an equal peak-to-peak signal range, the amplitude of noise required for an error decision is reduced to one-third that of the binary case.

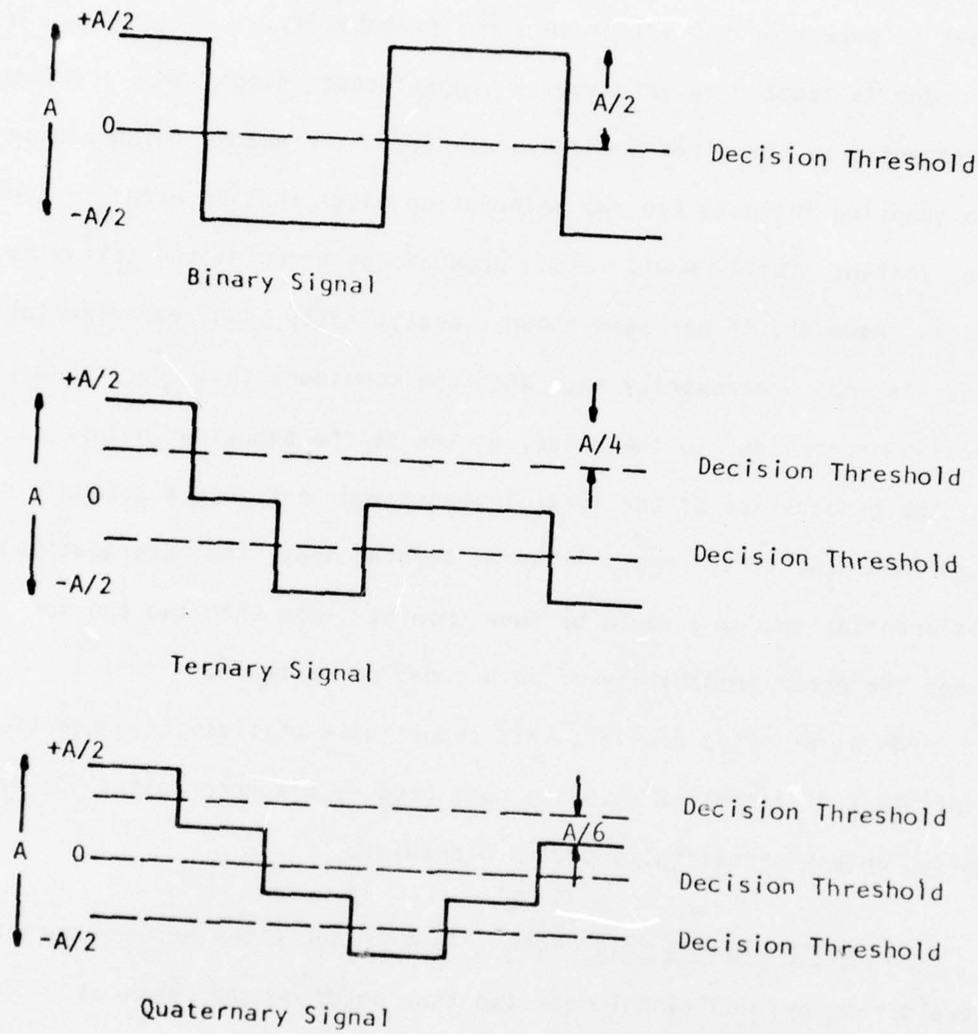


Figure 4.6. M-ary Polar NRZ Signal Decision Thresholds

In general, an m-ary signal format has a margin against error of 1/(m-1) that of the binary case. In the quaternary case, the two inner levels can be disturbed by either polarity of noise and are twice as likely to produce errors. In odd or even number of level systems, with symmetrical distribution around zero, the separation between adjacent levels is

$$D = A/(m-1) \tag{4-25}$$

where D is the distance in terms of the peak-to-peak signal range A.

An error will occur whenever the absolute value of the noise wave, at the sampling instant, exceeds one-half of this distance except when the signal sample is at the two extreme levels of $\pm$ A/2. When the signal sample is at +A/2, an error occurs only when the noise sample is less than -A[2(m-1)]; and when the signal sample is at -A/2, only when the noise sample is greater than +A/[2(m-1)].

If all levels are equally probable, the probability of any one level is 1/m, and the probability of an error in any one signal can be written in terms of the general expression for probability of error as

$$P_e = \frac{m-1}{m} \left(1-\mathrm{erf}\sqrt{M_m}\right) \tag{4-26}$$

We can compute the average signal power as

$$\frac{(m+1)\, S_{p2}}{3(m-1)} \tag{4-27}$$

where S_{p2} is the value for the binary polar NRZ case and equals $A^2/4$.

M_m can be shown to be

$$\left[\frac{3}{m^2-1}\right]\left[\frac{S_{p2}}{N_{p2}}\right] \qquad (4\text{-}28)$$

where $\frac{S_{p2}}{N_{p2}}$ is the average total signal power-to-noise ratio for the Nyquist II binary case. This provides a comparison to the binary case on an equal bandwidth or signalling pulse rate.

However, it must be remembered that, in comparing m-ary signals with binary symbols, the m-ary signal contains ($\mathrm{Log}_2 m$) bits of information per pulse and hence, on an equal information bit rate basis, only $1/(\log_2 m)$ times as many pulses per second, or Hz of bandwidth in the Nyquist II case, are required.

4-4. INFORMATION CODES

In order to transmit discrete messages (data or written messages) correctly and unambiguously, each symbol appearing in the message must be uniquely characterized. The process whereby a sequence of discrete bits are used to identify a larger piece of information is termed coding. It is the sequence of "1" and "0" (in the case of binary waveforms) that represent the numbers in a computer, a message over a teleprinter service, etc.

In order to assure the reliable transfer of information, it is important to code each symbol, comprising the message, into a code word without ambiguity. The length, or the number of bits, necessary to represent a symbol depends on the total number of possible symbols appearing at the source. Some of the commonly used codes are presented.

4-4.1. BAUDOT CODE. One of the earliest, codes is the Baudot 5 bit teleprinter code. Five bits are used to represent each character. Accordingly, 32 (2^5) different characters can be coded. Figure 4-7 shows the Baudot 5

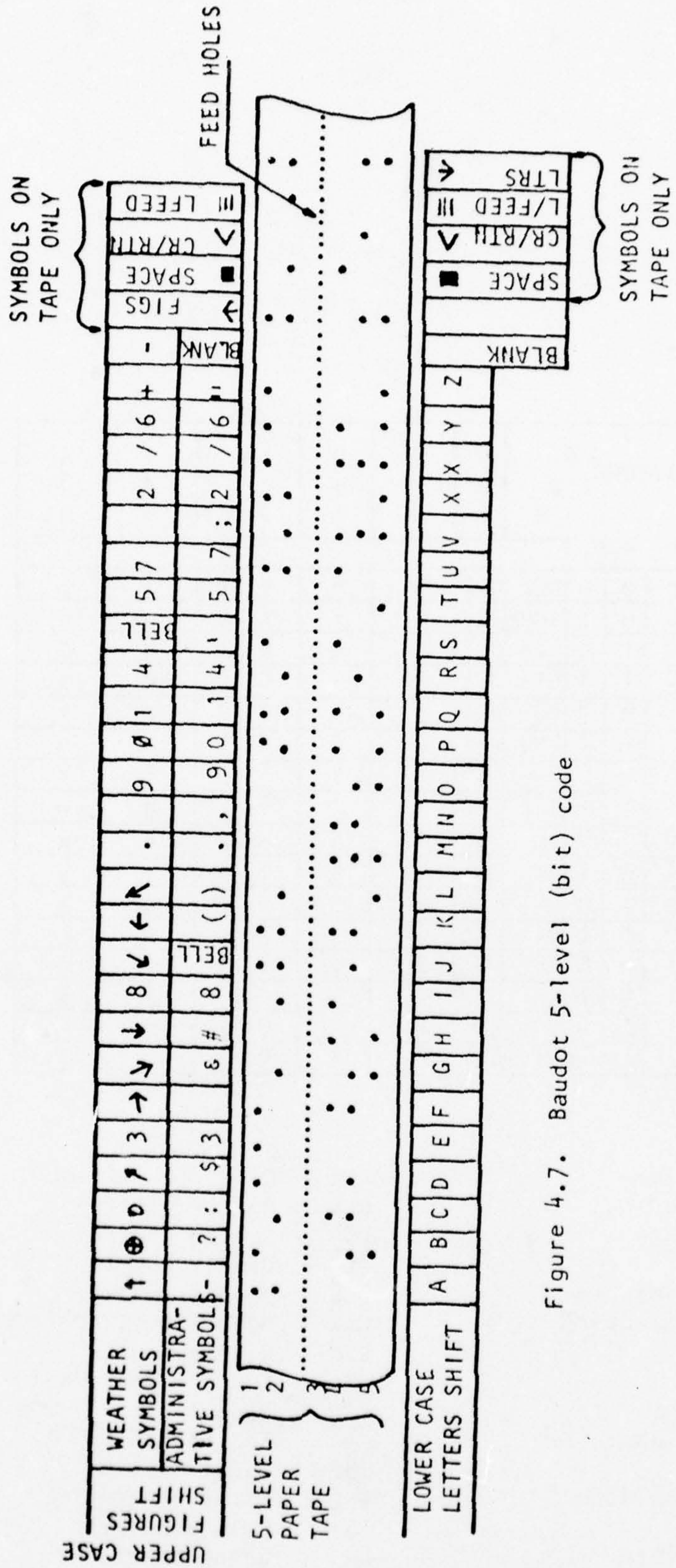


Figure 4.7. Baudot 5-level (bit) code

BIT POSITIONS			b_7 b_6 b_5 →	0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b_4	b_3	b_2	b_1								
0	0	0	0	NUL	DLE	SP	Ø	@	P	~	p
0	0	0	1	SOH	DC_1	!	1	A	Q	a	q
0	0	1	0	STX	DC_2	"	2	B	R	b	r
0	0	1	1	ETX	DC_3	#	3	C	S	c	s
0	1	0	0	EOT	DC_4	$	4	D	T	d	t
0	1	0	1	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	ACK	SYN	&	6	F	V	f	v
0	1	1	1	BEL	ETB	'	7	G	W	g	w
1	0	0	0	BS	CAN	(	8	H	X	h	x
1	0	0	1	HT	EM	)	9	I	Y	i	y
1	0	1	0	LF	SUB	*	:	J	Z	j	z
1	0	1	1	VT	ESC	+	;	K	[	k	{
1	1	0	0	FF	FS	,	<	L	\	l	¬
1	1	0	1	CR	GS	-	=	M	]	m	}
1	1	1	0	SO	RS	.	>	N	^	n	/
1	1	1	1	SI	US	/	?	O	_	o	DEL

LEGEND:

NUL Negative Value
SOH Start of Heading
STX Start of Text
ETX End of Text
EOT End of Transmission
ENQ Enquiry (Character)
ACK Acknowledge
BEL Bell
BS Backspace
HT Horizontal Tabulation
LF Line Feed
VT Vertical Tabulation
FF Form Feed
CR Carriage Return
SO Shift Out
SI Shift In
SO Space

DLE Data Link Escape
DC_1 Device Control
DC_2 Device Control
DC_3 Device Control
DC_4 Device Control
NAK Negative Acknowledge
SYN Synchronous
ETB End of Transmission Blcok
CAN Cancel
EM End of Medium
SUB Substitute
ESC Escape Character
FS File Separator
GS Group Separator
RS Records Separator
US Unit Separator
DEL Delete

Figure 4.8. 7-level ASCII software code

bit code. The number of distinct characters is increased by having an upper- and lower-case form. These two forms are indicated by using the letter shift and the figure shift keys on the teleprinter. Once the letter shift is sent, all the characters that follow are in lower case unless a figure shift is sent. Baudot code with few modifications is also referred to as International Telegraph Alphabet 2 (ITA-2) code (American version).

4-4.2. ASCII CODE. In data transmission, particularly with computers, the 5 bit teleprinter code poses too many restrictions. A widely used code in computer systems is the ASCII code (American Standard Code for Information Interchange). The ASCII code is defined with 7-bits per character; therefore, it has 128 (2^7) distinct combination of marks, spaces, etc. for assignment as characters. Users of the ASCII code may also represent each character by 8 bits, where the eighth bit is a parity bit, which is discussed later. The ASCII code is shown in Fig. 4-8.

4-4.3. BINARY CODED DECIMAL (BCD). (Figure 4-9). Data transmission frequently involves the transfer of numbers only, particularly in computer communications. These numbers may get very small or very large. Rather than converting them into the ASCII code, it is convenient to transmit the numbers in their binary form. Each digit of the number is transmitted sequentially in its binary representation. It takes four bits to represent the decimal digits 0,1,2,...,9. When each digit is converted to binary and transmitted, the resulting code is called Binary Coded Decimal. For example, the decimal number 26 in binary representation is 11010; in BCD it is 0010 0110 - the binary representation of digits 2 and 6 are combined to form the BCD representation. By assigning each alphabet character a unique binary coded decimal, this code can provide both alphabet and numeric

BIT				→ b1	0	0	1	1
POSITIONS				→ b2	0	1	0	1
b6	b5	b4	b3					
0	0	0	0		SOH	&	-	0
1	0	0	0		A	J	/	1
0	1	0	0		B	K	S	2
1	1	0	0		C	L	T	3
0	0	1	0		D	M	U	4
1	0	1	0		E	N	V	5
0	1	1	0		F	O	W	6
1	1	1	0		G	P	X	7
0	0	0	1		H	Q	Y	8
1	0	0	1		I	R	Z	9
0	1	0	1		STX	SP	ESC	SYN
1	1	0	1		.	$	,	#
0	0	1	1		<	*	%	@
1	0	1	1		BEL	US	ENQ	NAK
0	1	1	1		SUB	EOT	ETX	EM
1	1	1	1		ETB	DLE	HT	DEL

With BCD, 64 code combinations are possible to represent Alpha (26), Numeric (10), Symbols (11), and Control Functions (17) to include BELL (BEL) and SPACE (SP). By adding two bits to this code, it becomes 'Extended Binary Coded Decimal Inter-Change Code' (EBCDIC) with 256 total code combinations.

LEGEND:

SOH	Start of Heading	DLE	Data Link Escape
STX	Start of Text	ESC	Escape
BEL	Bell	ENQ	Enquiry
SUB	Substitute	ETX	End of Text
ETB	End of Transmission Blcok	HT	Horizontal Tabulation
SP	Space	NAK	Negative Acknowledge
US	Unit Separator	EM	End of Medium
EOT	End of Transmission	DEL	Delete
		SYN	Synchronous

Figure 4.9. 6-bit Binary Coded Decimal (BCD) transcode

BIT POSITIONS b1 b2 b3 b4 / b8 b7 b6 b5	0 0 0 0	0 0 0 1	0 0 1 0	0 0 1 1	0 1 0 0	0 1 0 1	0 1 1 0	0 1 1 1	1 0 0 0	1 0 0 1	1 0 1 0	1 0 1 1	1 1 0 0	1 1 0 1	1 1 1 0	1 1 1 1
0 0 0 0	NUL	DLE	DS		SP	&	-									0
1 0 0 0	SOH	DC_1	SOS				/		a	j			A	J		1
0 1 0 0	STX	DC_2	FS	SYN					b	k	s		B	K	S	2
1 1 0 0	ETX	DC_3							c	l	t		C	L	T	3
0 0 1 0	PF	RES	BYP	PN					d	m	u		D	M	U	4
1 0 1 0	HT	NL	LF	RS					e	n	v		E	N	V	5
0 1 1 0	LC	BS	ETB	UC					f	o	w		F	O	W	6
1 1 1 0	DEL	IL	ESC	EOT					g	p	x		G	P	X	7
0 0 0 1		CAN							h	q	y		H	Q	Y	8
1 0 0 1		EM							i	r	z		I	R	Z	9
0 1 0 1	SMM	CC	SM		¢	!		:								
1 1 0 1	VT				.	$	,	#								
0 0 1 1	FF	IFS		DC_4	<	*	%	@								
1 0 1 1	CR	IGS	ENQ	NAK	(	)	-	'								
0 1 1 1	SO	IRS	ACK		+	;	>	=								
1 1 1 1	SI	IUS	BEL	SUB	|	¬	?	"								

LEGEND:

NUL	Null
PF	Punch Off
HT	Horizontal Tabulation
LC	Lower Case
DEL	Delete
RES	Restore
NL	New Line
BS	Backspace
IL	Idle
CC	Cursor Control
DS	Digit Select
SOS	Start of Significance
FS	Field Separator
BYP	By Pass
LF	Line Feed
EOB/ETB	End of Block/or End of Transmission Block
PRE	Prefix
SM	Set Mode
PN	Punch On
RS	Reader Stop
UC	Upper Case
EOT	End of Transmission
SO	Space
SOH	Start of Heading
STX	Start of Text
ETX	End of Text
ACK	Acknowledge
BEL	Bell
VT	Vertical Tabulation
FF	Form Feed
CR	Carriage Return
SO	Shift Out
ENQ	Enquiry
SI	Shift In
SMM	Start of Manual Message
DLE	Data Link Escape
DC_1	Device Control
DC_2	Device Control
DC_3	Device Control
DC_4	Device Control
NAK	Negative Acknowledge
SYN	Synchronous
CAN	Cancel
EM	End of Medium
SUB	Substitute
IGS	Info. Group Separator
IRS	Info. Record Separator
IUS	Info. Unit Separator
IFS	Info. Field Separator

Figure 4.10. Extended Binary Coded Decimal Interchange Code (EBCDIC)

transfer.

4-4.4. EBCDIC. EBCDIC is an acronym for Extended Binary Coded Decimal Interchange Code. This code represents each character by one eight-bit number. These eight bits are divided into two groups of 4 bits each. Since 4 bits can represent the integers from 0 to 15, the numbers 10, 11, 12, 13, 14, 15 are represented by A, B, C, D, E and F. This forms a true hexadecimal code. The complete EBCDIC code is shown in Fig. 4-10.

4-4.5. HOLLERITH. The Hollerith Code is most commonly used on punched cards for reading input information to the computer. A character is represented on a card by one or more holes punched in one of the 12 hole positions on a card. Thus, Hollerith is a 12-unit character code. The number of symbols that can be represented by a 12-unit code is very large. Therefore, in common usage, no more than 3 holes are used in any column. The Hollerith Code is rarely used for transmission purposes; it merely functions to input data to the computer. Other codes, such as Field Data, EBCDIC, ASCII, etc. are used for transmission.

4-4.6. FIELD DATA CODE. (Figure 4-11). The Field Data Code is an 8-bit code which was used by the Military, but has been replaced by ASCII. Out of the 8-bits, 7-bits are used as information bits and the eighth bit is used for control. Thus, with the 7-bits allocated for information, 128 (2^7) bit patterns are available for assignment to characters. The 128 bit patterns are subdivided into 64 each. The control bit, along with the bit patterns, from each subdivision, is used for coding information. Due to the presence of the control bit, the Field Data Code has parity check capability which is discussed later.

FIELDATA 7-level (bit) plus 1 parity bit code is used as a standard military code in some digital communication systems. Out of a total of 128 code combinations, only 94 are used to represent either an alpha, numeric, symbol, or control function. Some code combinations may have been altered by extending or deleting some control functions or symbols to conform to system requirements.

BIT POSITIONS → b7 / b6 / b5				0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b4	b3	b2	b1								
0	0	0	0				TEL	BLANK	K	)	Ø
0	0	0	1			#	ACK-1	UC	L	-	1
0	0	1	0				ACK-2	LC	M	+	2
0	0	1	1			OWD	REQ	LF	N	<	3
0	1	0	0				WBT	CR	O	=	4
0	1	0	1			@	REP	SP	P	>	5
0	1	1	0			%	SOM-L	A	Q	—	6
0	1	1	1			¢	ER	B	R	$	7
1	0	0	0			BELL	DM	C	S	*	8
1	0	0	1	IL		&	EOM	D	T	(	9
1	0	1	0			MC	SOLB	E	U	"	'
1	0	1	1			≠	EDB	F	V	:	;
1	1	0	0			≢	EOLB	G	W	?	/
1	1	0	1			o	RM	H	X	!	.
1	1	1	0				SOM-H	I	Y	,	▱
1	1	1	1		TAB			J	Z	⊕	†

LEGEND:

IL	Idle Line	SOLB	Start of Line Block
TEL	Teletype Assignment	EDB	End of Data Block
OWD	One Way Delete	EOLB	End of Line Block
MC	Mode Change	RM	Reject Message
ACK-1,2	Acknowledge	SOM-H	Start of Message-High
REQ	Request	UC	Upper Case
WBT	Wait Before Transmitting	LC	Lower Case
REP	Repeat	LF	Line Feed
SOM-L	Start of Message Low	CR	Carriage Return
ER	Error	SP	Space
DM	Disregard Message	TAB	Tabulation
EOM	End of Message		

Figure 4.11. 7-level FIELDATA (military) code

4-5. ERROR DETECTION AND CORRECTION.

The reliability of transmission in digital systems is measured in terms of bit error probability. Since noise is present in all practical systems, there is a finite chance that a received bit in a bit stream is in error. The process of identifying an error in a received symbol is termed error detection. Consider, for example, a source that produces two possible messages. Then, provided the transmitter and receiver know beforehand the contents of the two messages, transmission can be accomplished as follows: A binary 0 is sent whenever the first message is produced and a binary 1 is sent whenever the second message is produced. Owing to noise - especially impulse noise - there will be errors. A "0" may be received as a "1" or vice versa, thus resulting in an erroneous received message. Furthermore, there is no way of telling if an error has occurred. A second way of transmitting the possible messages would be to represent the first message by two 0's and the second message by two 1's. Now, if a 01 or a 10 is received, an error has occurred; which of the two messages was sent is not known. In this case, however, the occurrence of a single error has been detected. Of course, there is a possibility of two errors occurring; that is, a 00 is received as a 11, or vice versa. In this case, again, there is no way of detecting two errors. The transmitted code could be made more complicated by using three binary digits to represent each message: a 000 indicates the first message; a 111 indicates the second message. If a 001, 010, or 100 is received, then there is an error. However, a decision rule can be established to correct such errors. In this case, since there are more 0s, the receiver can identify it as the first message. Similarly, if 110, 101, or 011 is received, the received message can be identified as the second one since there are more 1s.

The above example illustrates how a code can be made more complicated to combat the effect of noise. In the first method, when only a single digit was transmitted, there was no simple way to detect any errors. In the second method, single errors could be detected. In the third method, single errors could be detected and corrected.

The price paid for having error detecting and correcting capability through the use of simple bit redundancy in a code results in either a reduced rate of information transfer or "throughput" in the system or an increase in the bit rate to maintain the desired throughput. Three bits can be used to transmit eight different possible messages, but, in the above example, three bits are used to give single error detecting and correcting capability to a code used for representing two different messages. The transmission of more information than is strictly necessary for communicating all messages from a source in order to guard against errors (both due to noise and humans) is termed redundancy. Theoretically, error-free transmission is possible by adding enough redundancy in a code. This may, however, result in an extremely low information transfer rate, which may not be a practical solution to the error problem. The amount of redundancy used in a code depends on whether single, double, triple, or more errors are liable to occur more frequently in a transmission environment.

4-5.1. PARITY CHECKS. The bits used to represent a symbol or message produced by a source are called information bits. In order to have error detecting capability, a type of redundancy bits, called parity bits, can be added to the information bits to form a complete code word. A single parity bit may be added so that the sum of binary 1s in a code word is an odd number (odd parity) or an even number (even parity). Such an arrangement detects single bit errors; that is, a 1 changed to a 0 or vice versa can be

detected by single parity bit. The ASCII code is an example where a single parity bit is added to allow for error detection. The single digit parity check informs the receiver of a single error; it does not identify which digit is the wrong one; rather a request for retransmission is made whenever any error is detected.

Often code words are grouped into blocks, and a further special character is added as a block parity check. In such cases, the character parity check is referred to as the vertical parity check (VRC) and the block parity check is called the longitudinal parity check (LRC). In the VRC/LRC scheme, each data character to be transmitted is first checked for parity by the transmitting device, a parity bit is added as required (odd or even parity), and the resulting code word is then added (modulo 2) to the contents of a register and simultaneously transmitted. Each bit of the LRC character register serves as the resultant parity check on all data bits occupying the same bit positions (including the VRC parity bit). At the end of the data block, the resultant LRC character (in the transmitting register) is transmitted to the receiver. At the receiver, each character is checked for parity and is added (modulo 2) to its LRC registrar. At the end of the message block, the LRC character received from the transmitter is compared with the LRC character in the receiver LRC register. This method can detect all error patterns involving an odd number of errors within at least one of the bit positions containing errors, or an odd number of errors within at least one of the data characters. Thus, an error pattern with an even number of errors in all bit positions containing errors and in all data characters in error, will not be detected.

4-5.2. ERROR CORRECTION. In order to have error correcting capability - in addition to error detection capability - in a code, either more redundancy

digits are deliberately affixed to the information digits or complex coding/decoding schemes are utilized. These methods of error correction, whereby correction is made at the receiver on the basis of some decision rule, are called Forward Error Correction, in contrast to error detection and subsequent request for retransmission by the receiver.

Error correcting codes are, in general, more complicated and can be less efficient (in terms of effective throughput rate) than the codes which have error detecting capability only. Multiple parity digits can be used in order to detect errors, locate them, and then correct them. A set of codes, called Hamming codes, use multiple parity checks for detecting errors and correcting some of them. As an example of a Hamming code, refer to the so-called 4-out-of-7 code shown in Figure 4-12. This code uses 4 bits for information and 3 for parity check. The code shown uses an even parity check. Parity digit 1 is used to check information digits 1, 2, and 3; parity digit 2 is used to check information digits 1, 2, and 4; and, parity digit 3 is used to check information digits 1, 3, and 4. Note that the sum of 1s in the parity digit and the information digits it checks is an even number. This code will detect up to two errors and correct single errors.

In practical systems, errors occur in groups or bursts widely separated in time. For much of the transmission time there will be no errors, so correcting capability is not used. But when they do occur, the errors are multiple. Accordingly, error correcting codes must be selectively used. More often, on common voice channel paths, error detecting codes with automatic retransmission capability are used. Error correcting codes are most useful on special systems in which real-time data is required and the error rate in the data must be lower than that normally produced by the transmission path.

Information digit no. 1 2 3 4	Parity 1. Checks 1,2,3	Parity 2. Checks 1,2,4	Parity 3. Checks 1,3,4	Total Character
0 0 0 0	0	0	0	0000000
0 0 0 1	0	1	1	0001011
0 0 1 0	1	0	1	0010101
0 0 1 1	1	1	0	0011110
0 1 0 0	1	1	0	0100110
0 1 0 1	1	0	1	0101101
0 1 1 0	0	1	1	0110011
0 1 1 1	0	0	0	0111000
1 0 0 0	1	1	1	1000111
1 0 0 1	1	0	0	1001100
1 0 1 0	0	1	0	1010010
1 0 1 1	0	0	1	1011001
1 1 0 0	0	0	1	1100001
1 1 0 1	0	1	0	1101010
1 1 1 0	1	0	0	1110100
1 1 1 1	1	1	1	1111111

Figure 4-12. 4/7 Code with even parity.

4-5.3. BUFFERS. When error detection (parity) schemes are used, a protocol is required between the transmitter and receiver; i.e., after a code word or data block is transmitted, the receiver notifies the transmitter whether or not the received word or block contained errors, and if so, the transmitter retransmits the word or data block. Most systems permit retransmission of a word or block up to some specific number of times (three is a common number) and if it cannot produce an error free word or block in that number of times, the transmission path is determined to be "out-of-service" and appropriate alarms are activated. This type of operation is commonly called an "ARQ" system.

In any ARQ system some form of storage is required in order to "hold" each transmitted word or block until the receiver has acknowledged receipt of the word or block correctly. This type of storage is called a buffer. In some cases, the message source may "act" as its own buffer; while in other cases, a separate block in the system is required. In transmitting from a paper tape, for example, the tape reader reverses to the beginning whenever re-transmission is requested. The same can be true for messages from magnetic tapes or disks. In keyboard devices or card-readers, a separate storage unit may constitute a buffer.

4-6. REGENERATIVE REPEATERS

Another method of reducing the received errors in a digital system is to insert regenerative repeaters in the transmission path at points at which the expected worst-case S/N is high enough to permit detection of the incoming signal with a lower error rate than is possible at the end receiving point.

The function of a regenerative repeater is to detect, re-time, and reshape the digital waveform. Due to this ability to regenerate the transmitted waveform at regularly spaced distances, digital transmission uses a noisy media more effectively than analog transmission. In analog transmission, the noise is amplified along with the signal.

The functional diagram of a regenerative repeater is shown in Figure 4-13. For purposes of illustration, it is assumed that the pulse train at the output of the previous regenerative repeater, or the transmitter, is a bipolar wave. The waveform is distorted by the characteristics of the media and corrupted by noise. The waveform at the input of the regenerative repeater appears as shown at point B, Figure 4-13. The function of the preamplifier and equalizer block is to facilitate the subsequent retiming and reshaping of the wave. The preamplifier accentuates the level of the wave, so that a decision, called the level decision, about the presence and polarity of a pulse can be made by the regenerator block. A thresholding operation is used to check for polarity and the presence of a pulse or no-pulse. The timing block facilitates the regenerator to retime the distorted input pulse train. It provides a signal to sample the equalized pulse where the signal-to-interference level is the maximum, to maintain proper pulse spacing, and to time the output signal of the regenerator at the proper bit rate.

The regenerated wave may depart from being an exact replica of the transmitted wave a point A, Figure 4-13, due to any of the following factors: the interference is sufficiently high at the input so that a wrong decision is made, spacing between pulses departs from its proper value to an extent which precludes timing recovery and thus causing pulse position jitter, the output pulse shapes are not identical to the transmitted wave.

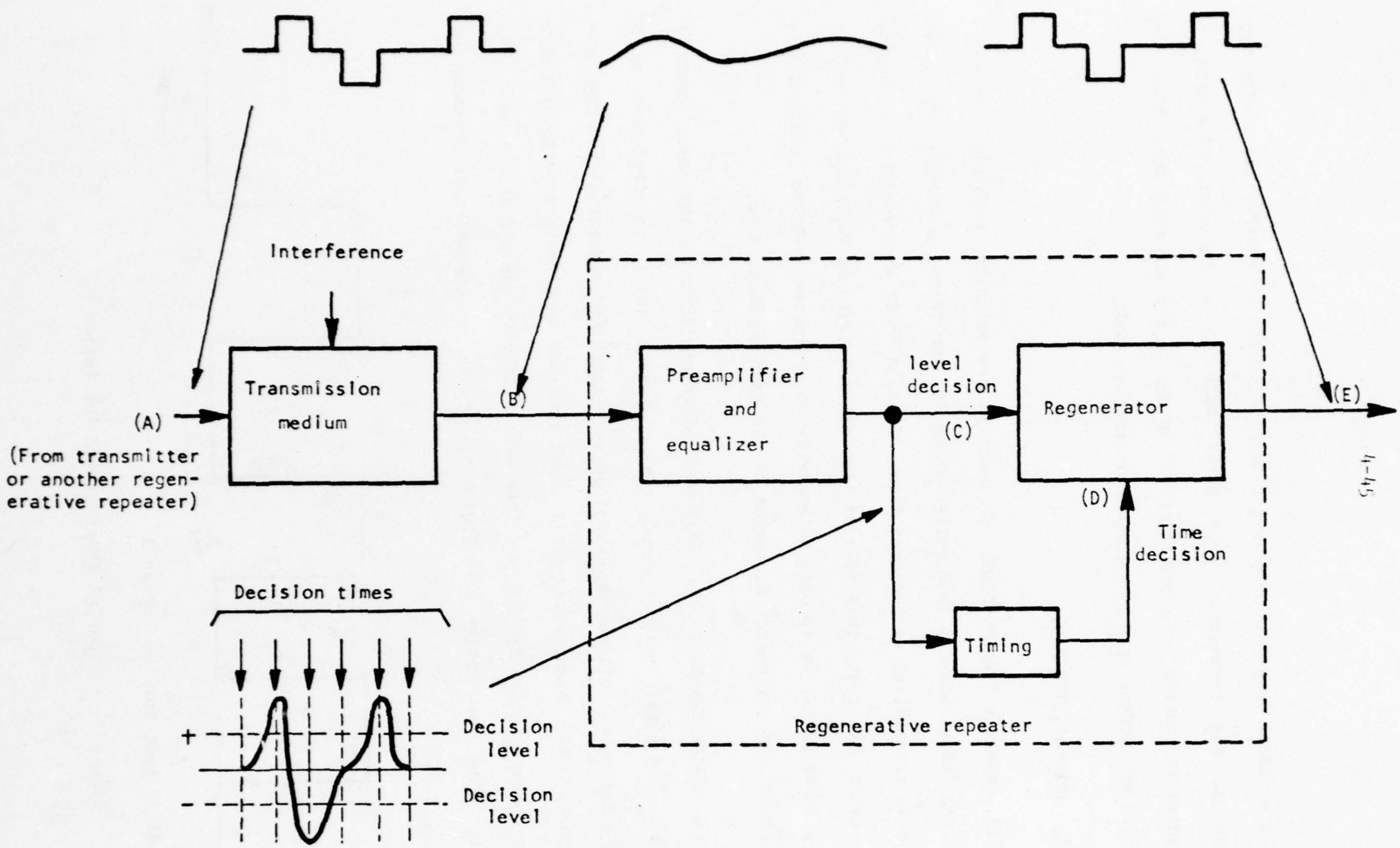


Figure 4-13. Regenerative repeater section block diagram.

Each of these imperfections have a bearing on the reconstruction of the message at the receiver. Each error made at a repeater, of course, is transmitted to the next repeater, and so on, to the end receiver; thus the errors are accumulated along the transmission path.

4-7. TRANSMISSION MODES

4-7.1. PARALLEL TRANSMISSION. In parallel transmission, each bit (element) of the code word is transmitted along its own channel, so that the entire word is transmitted at the same time. Thus, a 7-bit word needs 7 channels. In terms of cost, this mode of transmission would be expensive except for very short paths; it is used extensively in computer systems between the processor and peripheral equipment in the same physical area.

4-7.2. SERIAL TRANSMISSION. This mode of transmission is the most commonly used in digital transmission. Each element (bit) in the code word representing an information character is transmitted sequentially along one channel. For example, an ASCII coded terminal on-line to a computer codes each character into 8-bits, and then transmits them one bit at a time. The ASCII code for letter S is 01010011. Its serial transmission is shown in Figure 4-14.

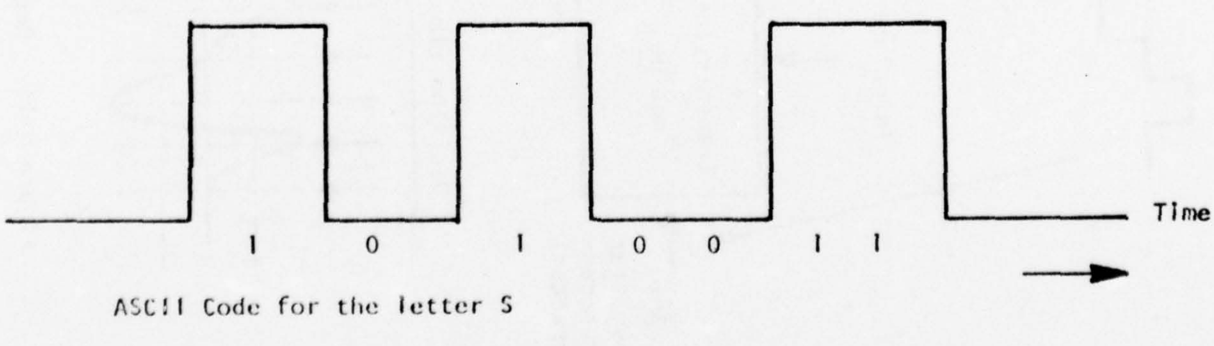


Figure 5-15. Serial Transmission for letter S.

4-8. TIMING/SYNCHRONIZATION

In order for the receiver to correctly decode the received code word, it must have information about where and when to expect the code for each word and for each bit in the word; i.e., it must be able to lock onto the transmitted signal, so that it can decode it. This is a timing problem. The concept of time is extremely important to digital transmission.

There are three basic methods of synchronization; symbol or bit timing, character or word timing and message or block timing. All systems must have some method of symbol or bit timing in order that the receive device can sample the incoming pulses and then compare the sample with its "Decision Threshold," to determine whether the pulse is a Mark or a Space (logic 1 or 0). Then, the system must employ either character (word) or block (message) timing in order to determine where a specific character or block of information starts and ends.

4-8.1. START/STOP-ASYNCHRONOUS SYSTEMS. In Telegraphy and many digital data systems, a method of timing called Asynchronous (Start-Stop) is widely used. This method combines character timing with bit timing. Each character code is preceeded by a single "Space" pulse (Start pulse) and followed by a "Mark" signal (Stop signal). Notice we said, "Mark signal," not "Mark pulse." This is because, depending on the design of the transmit and receive devices, the "Mark signal" may be equal to 1.0 bit, 1.42 bits, 1.5 bits, 2 bits, or any other time in length. The "Start" serves to phase the sampling of the code pulses and actually "Starts" the receive device detection or sampling mechanism. The "Stop" provides a pause or rest period before the start of the next character and assures that the receive device is ready to detect or sample the next character. Thus a "Stop" followed by a "Start"

signifies to the receiving device that a new character is beginning; and provides a reference point in time with which to judge where to sample each following pulse. The "Stop" signifies that the character is complete. Let's look at the Baudot code again (Fig. 4-15):

Table 4-1

Baudot Code Variation

Speed	Char Per Second	Length of Stop Pulse	Length of Start Pulse	Length of each of 5 Code Pulses
45.45 BPS	6.5	22 ms	22 ms	22 ms
45.45 BPS	6.13	31 ms	22 ms	22 ms
50 BPS	6.73	28.4 ms	20 ms	20 ms
50 BPS	6.67	30 ms	20 ms	20 ms
74.2 BPS	10.6	13.47 ms	13.47 ms	13.47 ms
74.2 BPS	10.0	19.18 ms	13.47 ms	13.47 ms
75 BPS	10.7	13.33 ms	13.33 ms	13.33 ms
75 BPS	10.0	19.99 ms	13.33 ms	13.33 ms

As shown in Table 4-1, the throughput, in terms of characters per second, can be different in two systems due to the difference in stop pulse length, even though the average bit-rates are the same for both systems.

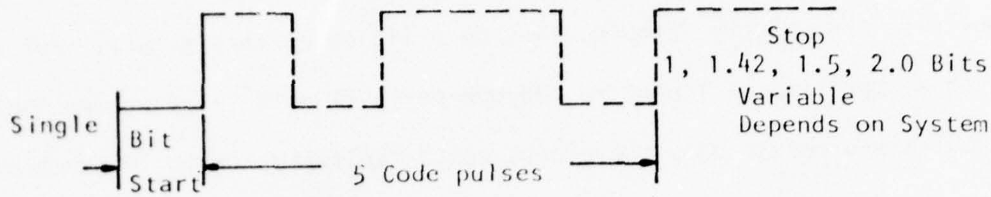


FIG. 4.-15
START-STOP FORMAT

4-8.2. SYNCHRONOUS SYSTEMS. The second widely used method of bit timing is called "synchronous." In this method the bit timing is actually recovered from the train of pulses which make up the signalling. There is no need for start or stop signals. The release of bits from the transmit device is usually precisely controlled by a highly stable oscillator referred to as a "clock." Each bit transmitted is then almost exactly the same length (typically to within one part in 100,000 parts). The receive device contains an equivalent oscillator (clock) and once in phase with the transmit oscillator will remain essentially in phase for reasonable periods of time, depending on the accuracy/stability of the two oscillator systems. The in-phase condition is also referred to as in "synchronization."

Synchronization is normally accomplished by transmitting a special pattern of pulses at the beginning of a block of characters or message. Upon recognition of the synchronizing pattern, the receive device brings its sampling timing in phase with that of the received signal and also because of the specific arrangement of one's and zero's in the pattern, knows exactly where it begins and ends and thus where the first bit in the first character will begin and end.

Although the oscillators (clocks) in the transmitting and receiving devices are highly stable, they do drift apart very slowly. For example, with a stability of 1 part in 100,000 parts (1×10^{-5}) at a sampling rate of 2500 times per second, the clocks would typically remain in synchronization for approximately 20 seconds. Re-synchronization must be periodically accomplished by inter-leaving the special synchronizing pattern with the message often enough so that the system never loses synchronization. This type of synchronous system is more properly referred to as being "Isochronous."

Another method of maintaining synchronization, which is truly synchronous, is to have a "controlled oscillator" or clock at the receive device. This type of clock gradually adjusts its timing rate to agree with that of the incoming signal and thus does not require periodic re-synchronization because of clock drifting or timing perturbations in the transmission path. It is widely stated in the literature that synchronous transmission is more efficient than asynchronous transmission because the start/stop bits are not required. Those making such statements, however, are confusing bit or symbol timing with character or block timing.

While it is true that start/stop pulses are not required in isochronous or synchronous systems to maintain <u>bit</u> timing, some system of identifying the beginning and/or ending of each character or block of data <u>is</u> required and these bits, referred to as framing bits, can be as extensive as the start/stop signals in asynchronous systems.

For instance, the ASCII code provides for a character start pulse equal in length to each information bit, and a stop pulse equal in length to one or two information bits. In a synchronous system, these start/stop pulses are not used in the receiver for bit timing, but only to determine where each character begins and ends.

In systems utilizing blocks of data, e.g., IBM cards, framing bits can be used to identify the beginning and ending of each block (card) and thus would be more efficient since a card, in this case, represents 80 characters of data.

In some systems, no character or block identification is used; i.e., a pattern is transmitted once (per message, minute, day, etc.) during each period of time and character/block synchronization is maintained through highly accurate synchronous bit timing with the character/block format known by both transmitter and receiver. This type of system is highly susceptible to loss of bit-count-integrity; i.e., if the transmission media inserts or deletes a bit due to noise, fading or other perturbations, this system loses synchronization and must be re-set. Synchronous cryptographic systems are a primary example of this type of system.

4-9. TRANSMISSION CIRCUIT MODES OF OPERATION

Another set of terms encountered in digital data transmission systems refer to the circuit mode of operation as "Simplex," "Half-Duplex," and "Duplex" or "Full-Duplex."

These terms refer to the operation of the circuit on an end-to-end basis; i.e., transmitter, transmission path, to distant receiver in both directions of transmission.

4-9.1. DUPLEX OR FULL-DUPLEX. This term refers to an end-to-end circuit in which both transmission path and terminal equipment are capable of transferring data in both directions simultaneously.

4-9.2. HALF-DUPLEX. This term refers to an end-to-end circuit in which the

terminal equipment can both send and receive, but not simultaneously. In most cases, the transmission path used is capable of two-way transmission simultaneously; this is so only because the long-haul trunk network (telephone) is inherently a 2-way network.

4-9.3. SIMPLEX. This term refers to an end-to-end circuit that permits transmission in one direction only; the terminals are commonly referred to as "transmit only" and "receive only."

4-10. DCS AUTOMATIC DIGITAL NETWORK (AUTODIN)

4-10.1. The AUTODIN is a store-and-forward message switched network consisting of large automatic switching centers interconnected by transmission trunks. Each subscriber normally homes on more than one switching center for survivability purposes. The AUTODIN can also dial-up AUTOVON paths for use as back-up trunks.

Although an exclusively data network, actual transmission is over voice-channel paths using modems to condition the digital signals.

4-10.2. An AUTODIN terminal sends a message over an access circuit to its switching center; here the message is temporarily stored. The switching center processor determines the proper routing required to deliver the message to the addressee(s); it also converts the message (code and bit-rate) to the proper trunk format and forwards the message to the switch(es) serving the addressee(s). The receiving switch stores the message, determines the routing to the addressee(s) (another switch or a connected terminal), converts the message to the proper code and speed for the addressee terminal, and forwards it to the terminal.

As can be seen, there is no direct electrical path between the sending terminal (originating) and the receiving terminal (addressee); each complete message is stored and forwarded by the switching center(s).

4-10.3. AUTODIN provides for five modes of operation.

MODE I - Full-duplex synchronous with automatic error detection, ARQ and channel control protocol.

MODE II - Full-duplex asynchronous without error detection, ARQ or channel control protocol.

MODE III - Hybrid half/full duplex synchronous with automatic error detection, ARQ and channel control protocol. Messages flow in one direction at a time; the return path is only used to send channel-control protocol and ARQ characters.

MODE IV - Simplex (one-direction only, not reversible) asynchronous without automatic error detection, ARQ or channel control protocol (usually a receiver only path switch to terminal).

MODE V - Duplex synchronous with limited channel control protocol.

The asynchronous modes are limited to 600 b/s and the synchronous modes to 4.8 kb/s.

4-10.4. The AUTODIN access circuit from the terminal (base telecom center) to the switching center is composed of two segments; the on-base segment which is a metallic path over the base cable plant, and the off-base segment which may be an individual metallic path, or a voice channel on a multiplexed wire and/or radio path.

Depending on the bit-rate, this access circuit must be conditioned to meet the applicable DCS specifications which are similar to those shown in Table 3-6.

As can be seen, there is no direct electrical path between the sending terminal (originating) and the receiving terminal (addressee); each complete message is stored and forwarded by the switching center(s).

4-10.3. AUTODIN provides for five modes of operation.

MODE I - Full-duplex synchronous with automatic error detection, ARQ and channel control protocol.

MODE II - Full-duplex asynchronous without error detection, ARQ or channel control protocol.

MODE III - Hybrid half/full duplex synchronous with automatic error detection, ARQ and channel control protocol. Messages flow in one direction at a time; the return path is only used to send channel-control protocol and ARQ characters.

MODE IV - Simplex (one-direction only, not reversible) asynchronous without automatic error detection, ARQ or channel control protocol (usually a receiver only path switch to terminal).

MODE V - Duplex synchronous with limited channel control protocol.

The asynchronous modes are limited to 600 b/s and the synchronous modes to 4.8 kb/s.

4-10.4. The AUTODIN access circuit from the terminal (base telecom center) to the switching center is composed of two segments; the on-base segment which is a metallic path over the base cable plant, and the off-base segment which may be an individual metallic path, or a voice channel on a multiplexed wire and/or radio path.

Depending on the bit-rate, this access circuit must be conditioned to meet the applicable DCS specifications which are similar to those shown in Table 3-6.

SECTION V

MODULATION SCHEMES FOR DIGITAL TRANSMISSION

5-1. GENERAL

Two primary methods can be utilized to transfer digital signals over base wire transmission paths; i.e., direct digital (baseband or modified baseband) and quasi-analog. Direct digital signalling can be used only if the transmission path is solely composed of metallic wire; i.e., there are no analog devices (amplifiers, loading coils, analog frequency division multiplex or analog radio equipment) in the path. Unless specific arrangements are made with the off-base carrier, all base wire paths which connect to the DCS or commercial common carrier networks should be assumed to contain analog devices and thus direct digital signalling cannot be utilized.

The basic reasons for this restriction are twofold; first, direct digital signals (except those based on the bipolar type of format) contain a DC component which cannot be passed over the existing analog network, and secondly, the existing analog network is constructed around a frequency division multiplexing hierarchy that provides transmission channels which are severely restricted in bandwidth. The inputs and outputs of these derived channels contain relatively sharp cut-off filters which preclude signalling outside of the assigned bandwidth. In addition, those channels were designed for the transmission of telephone voice signals, and although the frequency and phase distribution characteristics are adequate for this purpose, those same characteristics degrade data signals.

The primary channel provided by those networks is the nominal 4KHz voice channel; for transmission purposes, however, it has a useful bandpass of about 300-3200 Hz.

Direct digital signalling devices for bit rates above a few hundred bits-per-second are relatively new devices whereas modems for use over the voice channel are readily available. For this reason, most existing base-wire digital transmission paths utilize voice channel modems even though the base-wire path is not restricted to the voice channel bandwidth.

This section of the report presents the various modulation/demodulation schemes in common use for transmission of digital signals over voice channel bandwidths.

The techniques of direct digital signalling over base wire paths will be covered in a subsequent technical report.

For the purposes of this section of the report, modulation is defined as the alteration of an audio frequency sinusoidal carrier by a baseband pulse stream; demodulation is the reverse process.

The sinusoidal carrier signal may be altered (modulated) by varying its amplitude, frequency, or phase (and combinations thereof) in accordance with the state of the information bit stream. These modulation schemes are discussed and compared in this section of the report.

The theoretical performance of each scheme discussed, and the comparisons among the various schemes, will be based on the definitions and assumptions discussed in Section IV, paragraphs 4-3.6 and 4-3.7.

5-2. AMPLITUDE MODULATION

5-2.1. SINUSOIDAL AM. In the modulation of a sinusoidal carrier by a sinusoidal modulating wave (when the carrier frequency is much larger than the modulating frequency), 100% modulation is obtained when the modulating signal is so biased and limited in amplitude that its maximum positive direction amplitude causes the carrier wave to reach its maximum average am-

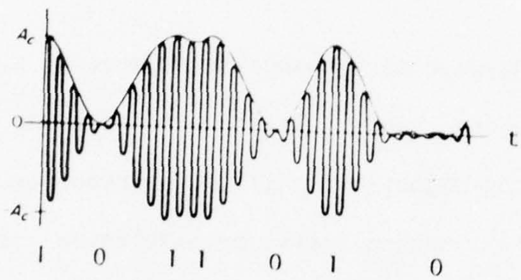


Figure 5-1. Binary Amplitude Shift Keying Waveform

plitude, its bias level crossing value causes the carrier wave to attain its one-half average amplitude; and its maximum negative direction amplitude causes the carrier wave to reach zero.

It can be shown that this form of modulation provides a useful composite signal which consists of the carrier frequency and a pair of frequencies, one located at a frequency equal to the carrier frequency plus the modulating frequency and one equal to the carrier frequency minus the modulating frequency.

If the modulating wave is a composite of several sinusoidal waves, the side frequencies become bands of frequencies equal in relative spectrum shape to the modulating signal but shifted in frequency.

Thus, 100% double sideband amplitude modulation is a linear process (when coherent detection is utilized).

If the modulating wave is more complicated but expressable by a series of sine and cosine terms, each term will lead to a pair of sideband frequencies.

It can also be shown that one-half of the total power is contained in the carrier wave and that one-fourth of the total power is contained in each of the sidebands.

With less than 100% modulation, the sideband power is proportional to the square of the modulation percentage.

5-2.2. SQUAREWAVE AM. *If a neutral baseband signal* (see Figures 5.1 and 5.2) of alternating marks and spaces (on-off pulses) is used to amplitude modulate a carrier, the DC component appears as the carrier term and the AC components appear as the sidebands with a spectrum similar to the baseband signal.

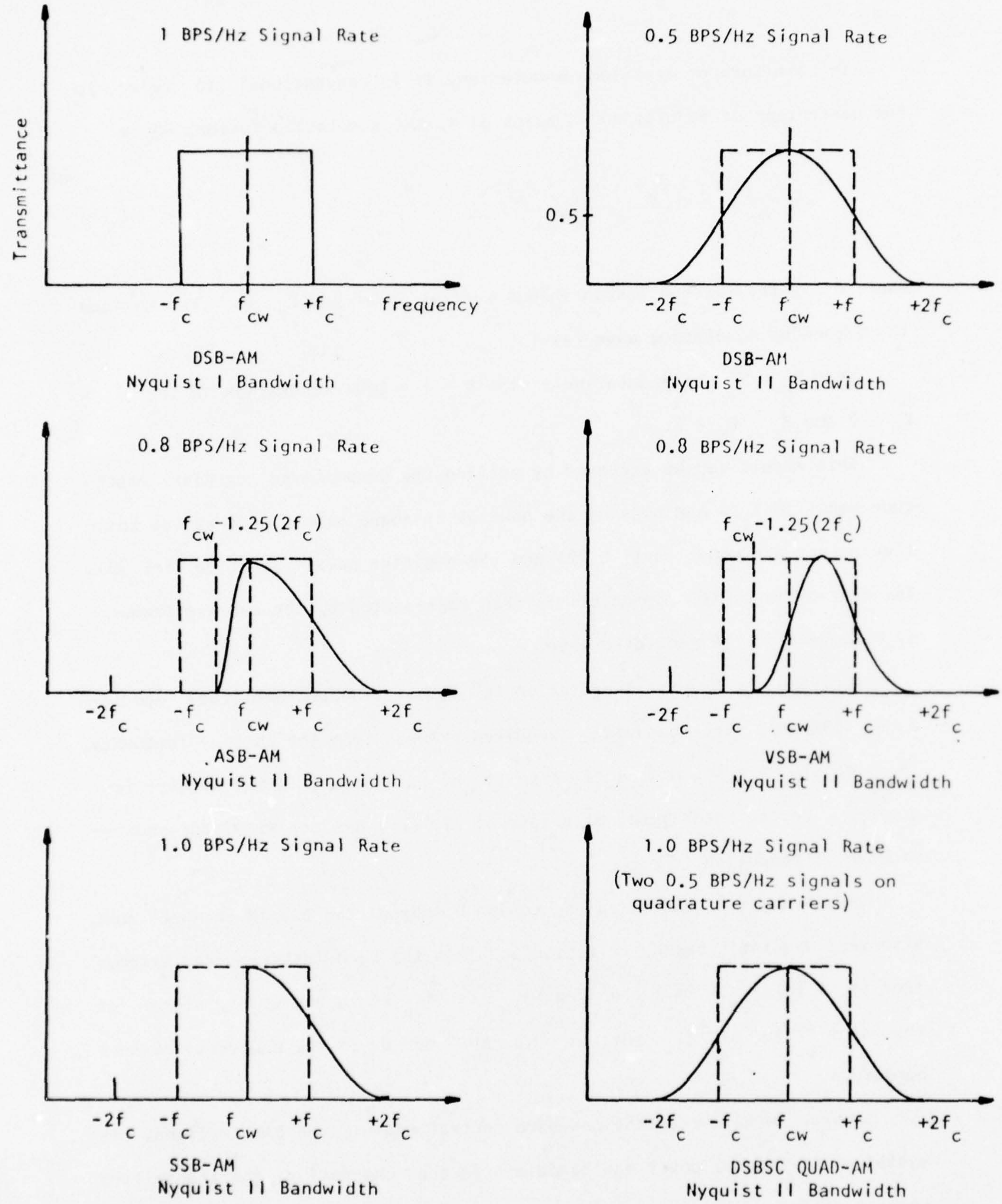


Figure 5.2. Amplitude-Modulated Signal Bandwidths

In this form of amplitude modulation, it is conventional to refer to the percentage of modulation in terms of K, the modulation factor, where

$$K = \frac{1}{2} \frac{[(E_c + E_m) - (E_c - E_m)]}{E_c} \tag{5.1}$$

where E_c is the maximum instantaneous carrier level and E_m is the maximum instantaneous modulating wave level.

When $E_m = E_c$ the modulation factor K = 1 = 100% modulation in that $E_c + E_m = 2$ and $E_c - E_m = 0$.

This result can be arranged by setting the unmodulated carrier amplitude equal to $E_c/2$ and biasing the neutral baseband signal so that the positive pulses are equal to $(+ E_c/2)$ and the negative pulses equal to $(-E_c/2)$. The biasing level (DC component) is then represented by the carrier frequency component in the modulated wave.

Applying the Nyquist I criterion to double-sideband-amplitude modulation (DSB-AM), the bandwidth required extends from the carrier frequency (f_{cw}) to $(f_{cw} + f_c)$ and from (f_{cw}) to $(f_{cw} - f_c)$, where f_c is the Nyquist frequency. We can then signal at a rate of $2f_c$ or 1 BPS per Hz of the double-sidebanded bandwidth.

On a Nyquist II basis (raised cosine baseband) the DSB-AM channel must also meet Nyquist's second criterion and thus the bandwidth required extends from (f_{cw}) to $(f_{cw} + 2f_c)$ and from (f_{cw}) to $(f_{cw} - 2f_c)$, and we now signal at the same rate of $2f_c$ but at 0.5 BPS per Hz of the double-sidebanded bandwidth.

Similar in nature to the baseband neutral signal, the DSB-AM signal is wasteful of signal power and bandwidth in that one-half of the transmitted power contains no information other than the biasing level DC component and

both of the sidebands contain identical information. The signal can be detected on an envelope detection or coherent basis with a cost of 0.5 dB in S/N sensitivity for the former.

Since contributions from both sidebands add to form the recovered baseband signal, arbitrary modifications can be made in the transmission characteristics of the two corresponding frequency ranges.

One modification, asymmetrical sideband (ASB-AM) transmits one complete sideband and a portion of the other sideband and typically requires about 1.25 times the bandwidth of the baseband signal compared with 2.0 times for DSB-AM. The lowest baseband frequencies are transmitted by DSB and the higher baseband frequencies by single side band (SSB). The amplitude of the lower frequencies is double that of the higher frequencies. The complete spectrum shows a 6 dB drop in the transition region from full DSB to complete SSB and usually an equalizer is inserted in the system to perform a reciprocal function.

Another modification, which is complementary, is vestigial sideband amplitude modulation (VSB-AM), in which partial suppression of one sideband in the neighborhood of the carrier is exactly compensated by partial transmission of the corresponding part of the other sideband. This technique conserves bandwidth and signal power and the gradual transition in frequency, between full transmission of one sideband and complete suppression of the other sideband, is a realizable function with practical filters.

Another modification, single-sideband, suppressed carrier amplitude modulation (SSBSC-AM) conserves bandwidth and signal power to an even greater degree in that one sideband and the carrier are completely suppressed. Disadvantages or costs incurred are the absence of a DC component and suppression of the very low frequencies in a realizable system

and the requirement for coherent detection; it is also sensitive to phase errors in the receive signal since it contains a quadrature phase component which contains as much energy as the in-phase component.

VSB-AM also contains a quadrature component, but it is not as large as in SSB-AM, and thus not as sensitive to phase errors at the detector.

Generation and detection of an SSB-AM signal is made easier if the baseband signal contains no DC component or very low frequencies; then an initially generated DSB-AM signal would also lack the DC component (carrier) and the sidebands would be separated in the region of the carrier frequency making suppression of the one sideband realizable with practical filters.

Baseband signals which do not contain a DC component are polar in nature (zero bias level) and those containing only small low-frequency components are bi-polar in nature. By our definition of 100% modulation (K = 1.0) the baseband signal was biased so that the negative going peak caused the carrier to be zero. With an unbiased baseband signal, the zero-level or crossing would cause the carrier to be zero and the negative peak of the baseband polar signal would be cut-off in an off/on amplitude modulator.

However, if we arrange the modulator so that the negative peak of the baseband polar signal causes the carrier to reach the identical level as the positive peak, but with an 180° phaseshift (K = 2.0), the result is a double-sideband signal, without the carrier component, in which the baseband marks and spaces are of equal amplitude in the modulated waveform, but displaced in phase by 180°.

This signal can be transmitted in this form as a double-sideband suppressed carrier (DSBSC-AM) signal or filtered to provide a SSBSC-AM signal.

The DSBSC signal requires the same bandwidth as the DSB-AM signal but provides a power gain since the carrier component is suppressed.

In the coherent detection of DSBSC-AM, it can be shown that with a 90^{o} phase error in the local carrier phase, the recovered signal terms vanish. This shows that we may transmit and receive two independent DSBSC-AM waves in the same signal band with both signals having the same carrier frequency, but different in phase by 90^{o}. This type of signaling is referred to as DSBSC-QUAD-AM or simply QUAD-AM. On this basis we now enjoy a bandwidth efficiency equal to that of the SSBSC-AM signal.

5-2.3. COMPARISON OF AM SIGNALLING (Refer to Fig. 5-3). The probability of error for amplitude modulated signals depends, of course, on which form is utilized and also depends on the type of detector [envelope (where applicable) or coherent], modulation index, and, in the cases of ASB and VSB, on the filter shaping and its affect on the level of the quadrature component in relation to the in-phase component and thus the sensitivity to phase error. In the case of SSB and Quad-AM, the signal has maximum sensitivity to phase error. For comparison purposes we shall assume ideal coherent detection.

It can be shown that m-ary SSBSC-AM and Quad-AM are equivalent to m-ary polar baseband signaling (P_e vs. S/N).

If the VSB-AM scheme is modified to VSBSC-AM scheme, this method also produces equal performance to its equivalent baseband signals and to SSBSC-AM.

In SSBSC-AM and VSBSC-AM, with coherent detection, one-half of the signal power is contained in the quadrature component, which is not utilized by the detector. However, in DSBSC-AM the signal power in the two full-sidebands adds linearly in the detectors so that on an equal AM wave power

m = NR levels m-level
mp = NR levels partial response

curve (a) Polar NRZ baseband, VSBSC, SSBSC, QUAD-AM on average total signal power constraint

(b) Same as (a) under maximum steady-state power limit

(c) Polar NRZ baseband VSBSC, SSBSC, QUAD-AM modulating FM radio with Mod. Index = 0.167 on max. steady-state power limit

(d) Partial response baseband, VSBSC, SSBSC, QUAD-AM on max. steady-state power limit

(e) Partial response baseband, VSBSC, SSBSC, QUAD-AM modulating FM radio with mod. index = 0.167 on max. steady-state power limit

NOTE:

(1) Curves (a), (b), and (d) also apply to baseband/AM/AM RF radio

(2) DSBSC-AM 3 db better than all curves at one-half capacity BPS/Hz

(3) For duobinary, mp = 3 points on curves (d) and (e) are applicable at $Pe=2\times10^{-4}$

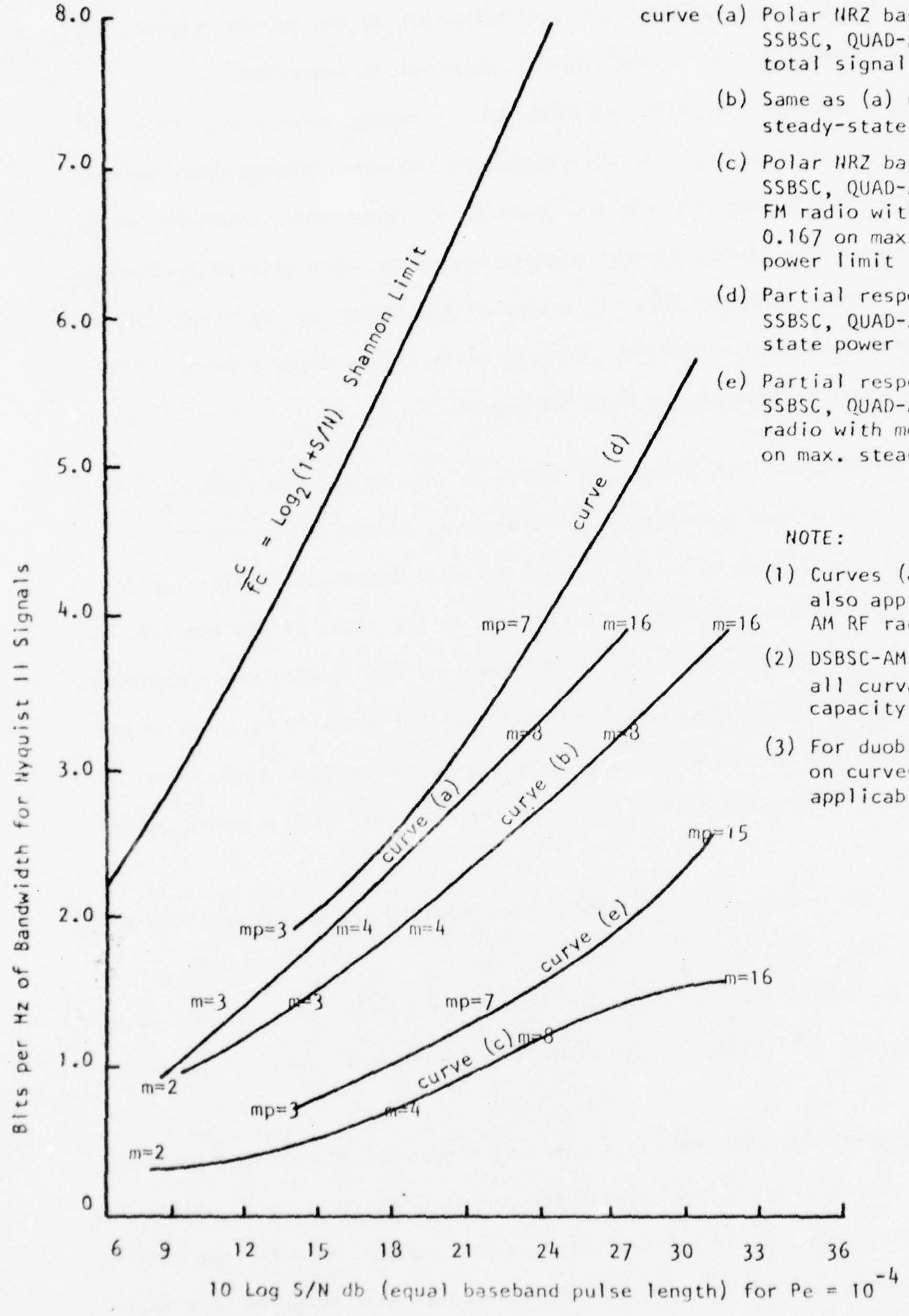


Figure 5.3. Bandwidth Efficiency in Amplitude Modulation

and bandwidth basis, DSBSC-AM has a 3 dB advantage over its equivalent baseband signal. On an equal information bit rate basis, the 3 dB advantage is cancelled by the 3 dB increase in noise power in the double bandwidth required.

DSB-AM, on a binary basis with equal average total power, suffers a 3 dB disadvantage to DSBSC-AM since one-half of the power is contained in the carrier component. This loss diverges rapidly as the number of levels is increased in m-ary signaling. On a maximum steady-state power limit basis, DSB-AM (ON-OFF) suffers an additional 3 dB loss (binary) since the average signal power is then only one-half that of the DSBSC-AM signal.

On a Nyquist I bandwidth basis, SSBSC-AM, VSBSC-AM and Quad-AM using 8-level signals approach 75% of the Shannon limit. If the Nyquist II bandwidth is considered the 8-level signal performance, of course, approaches only 37.5% of the Shannon limit.

5-3. FREQUENCY MODULATION (FM)

5-3.1. SINUSOIDAL FM. Frequency modulation is accomplished by the process of altering the frequency of the carrier wave in accordance with the amplitude of the message signal. The amplitude of the frequency modulated waveform is the same as that of the carrier; but its frequency at any instant is generally not the same as that of the carrier. The difference

between the instantaneous frequency of the FM signal and the carrier frequency is defined as the instantaneous frequency deviation. Under the frequency modulation scheme, it is the instantaneous frequency deviation that is proportional to the message. The maximum deviation in frequency from that of the carrier during the transmission of a message is called the peak frequency deviation. The bandwidth required for most practical (low-index) FM systems is approximately two times the sum of the highest frequency in the modulating signal (message) and the peak frequency deviation. In actual implementation, the bandwidth requirement for frequency modulated signals is dependent on the message and the desired quality of transmission.

a. An FM system with a wide swing has a signal-to-noise advantage over a comparable double-sideband AM system. The improvement, expressed as a ratio, is theoretically the deviation ratio times the square root of three. This relation holds only when the peaks of noise at the receiver are less than about one-half the carrier amplitude. Above this, the noise advantage of the FM system rapidly disappears, soon reaching a condition when the system output is substantially all noise. The condition is known as breaking of the system. Breaking occurs for more or less steady kinds of noise. Noise of the impulse type can have higher peak amplitudes before causing serious deterioration. Generally, a wide-swing system, because of the wider bandwidth of the receiver and the larger amount of noise it accepts, will reach breaking sooner than a narrow-swing system with a narrow band receiver, even though the wide-swing circuit has a higher quality in the noise amplitude range.

b. Noise in an FM receiver tends to have a triangular spectrum at the demodulated output. That is, with equal amplitudes of the noise components at the input to the receiver, the demodulated outputs show progressively

larger amplitudes at the higher frequencies. However, at these high frequencies, where the noise is strong, most types of signals (such as speech) have less energy. A further improvement in signal-to-noise ratio can therefore be obtained in such cases by introducing networks in the transmitter to emphasize the high-frequency signal components so that they can better override the noise. The signal is restored to normal by adding compensating deemphasizing networks in the receiver.

c. Interference in an FM receiver from an unwanted FM transmitter, operating on the same or on a closely adjacent frequency as the desired transmitter, also has a unique effect. The interference effect of the unwanted transmitter (for a deviation ratio of five or more) is quite small if its field (at the frequency being received) is less than about one-half as strong as the desired field, and practically disappears when it is about one-third as strong. Increasing the field strength of the interfering transmitter rapidly introduces a garbling of the signal so that neither station is understandable. Further increase of the interfering carrier to two or three times that of the desired station permits the interfering signal to come through clearly with little evidence of the wanted signal. The almost complete taking over of a receiver by the appreciably stronger of the two FM transmitter fields is called the capture effect.

d. The wider the swing of an FM system, the more frequency space it occupies. On the otherhand, a wider swing improves the noise advantage of an FM system. Therefore, it is evident that the additional noise advantage is obtained at the cost of extra required bandwidth. This characteristic, in which bandwidth can be traded for noise advantage, is a common one, and

is called wideband noise reduction. It is found in other modulation schemes, too. It may be an important consideration in selecting or engineering a transmission system, especially one in which bandwidths are available, but power is limited.

5-3.2. FREQUENCY-SHIFT KEYED (FSK) SIGNALING. FSK is a form of frequency modulation. Here, we are addressing those signal processing schemes in which the digital stream is converted to an FSK signal. Analysis of FSK schemes is not as straightforward as baseband and AM schemes in that the capacity, bandwidth required, signal spectrum, and performance obtained, while functions of the deviation ratio and the method of signal generation and detection, are not straightforward power and bandwidth conversions from one scheme to the next, as was the case in AM.

In binary FSK signaling, one state is represented by frequency f_1 and the opposite state by frequency f_2 (see Figure 5.4). The deviation ratio h is expressed as

$$h = (f_2 - f_1)T \tag{5-2}$$

where T is the bit interval length of the input signal.

Typical deviation ratios range from 0.5 to 16, with the latter value being typical of older HF radio narrow band channels. A rule for determining the required bandwidth for FSK emission is given by

F_1 = BK + 2D where

F_1 is the FSK required bandwidth,

B is the signaling speed in bauds,

D is one-half the difference between the maximum and minimum frequencies with D being greater than B, and

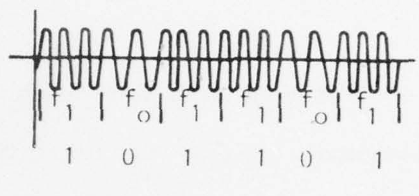


Figure 5.4. Binary Frequency Shift Keying Waveform

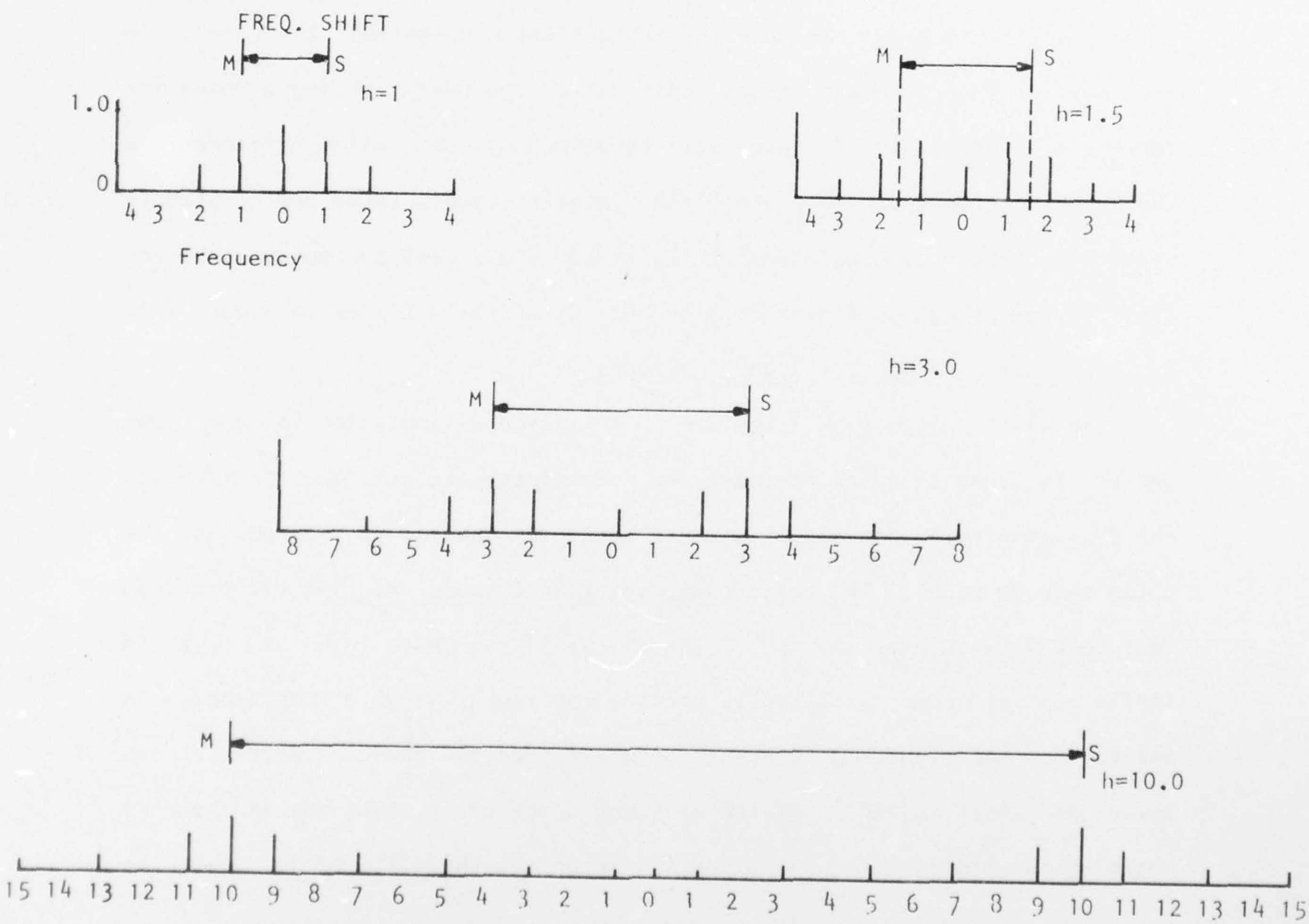


Figure 5.5. FSK SPECTRA

K is a constant which has a value of 3 for non-fading media and 5 for fading media.

Under this formula a 50 baud teletype channel utilizing FSK with an 800 Hz shift requires a 950 Hz bandwidth for a non-fading media and 1050 Hz for a fading media.

Bandwidth requirements of this magnitude are reasonable with large deviation ratio's when one considers that as the ratio becomes large, energy spreads away from the mid-frequency $(f_1 + f_2)/2$ and concentrates near the steady-state mark and space frequencies. For integer values of the deviation ratio, the result is the sum of the two on-off amplitude modulated carriers located at the f_1 and f_2 steady-state frequencies, and keyed in phase opposition. This occurs because the integer ratio preserves the phase in the carrier from one mark (space) interval to the next. During a space interval, a whole number of cycles will be gained or lost with reference to the mark frequency. For other than integer values, there are no steady-state side frequency components at the steady-state mark and space frequencies and both even and odd order sidebands are present (refer to Figure 5.5).

a. Application of Nyquist Signaling to FSK.

The system given most attention in this type of analysis is one suggested by Sunde in which the mark and space intervals are equal in time and the deviation ratio h = 1.0, i.e., the frequency shift in Hz equals the pulse rate in bauds. The total phase-shift in a space interval differs from that in a mark interval by 360°. Therefore, if the phase of the wave is continuous at one transition it will be continuous at all transitions. In one-half of the signaling interval the difference in phase change between space and mark is 180°. If the mark and space oscillators are 180° out of phase at the mid-point of the signaling interval, they will be in phase at

the beginning and end of the interval and conditions are thus established for continuity of the phase at the switching instants. Under these conditions we can apply the Nyquist criteria to FSK as shown in Figure 5-6.

From Figure 5-6 it appears that Sunde's system requires the same bandwidth as binary DSB-AM for the same pulse rate. However, this conclusion must be qualified by observing that while the instantaneous frequency has been controlled, no attempt has been made to obtain a constant amplitude wave, and hence the solution is not strictly an FM wave; it is a hybrid wave containing both AM and FM. A pure FM wave can be provided by limiting the amplitude but this will change the spectrum and in general will widen the range of required frequencies. The signaling information is in the FM wave, but some AM is necessary to fit the complete wave into the desired bandwidth. The wave consists of unmodulated sinewaves plus suppressed carrier AM.

In analog FM, it is convenient to have the deviation larger than the baseband bandwidth to provide S/N ratio improvement as a tradeoff of bandwidth for signal power. In binary FSK, if the bandwidth is sufficient to prevent intersymbol interference and the S/N ratio is sufficient, no errors occur and increasing deviation and thus bandwidth, would increase the noise power. For m-ary FSK systems with $m > 2$, one can trade-off, since the higher S/N ratio is required. As $m \to \infty$, the analog and digital signals approach equivalence.

b. Probability of Error

In the discussion of FM signalling we took cognizance of the triangular noise spectral density resulting from the non-linear FM demodulation process. In FSK the detector is also non-linear and the non-Gaussian noise

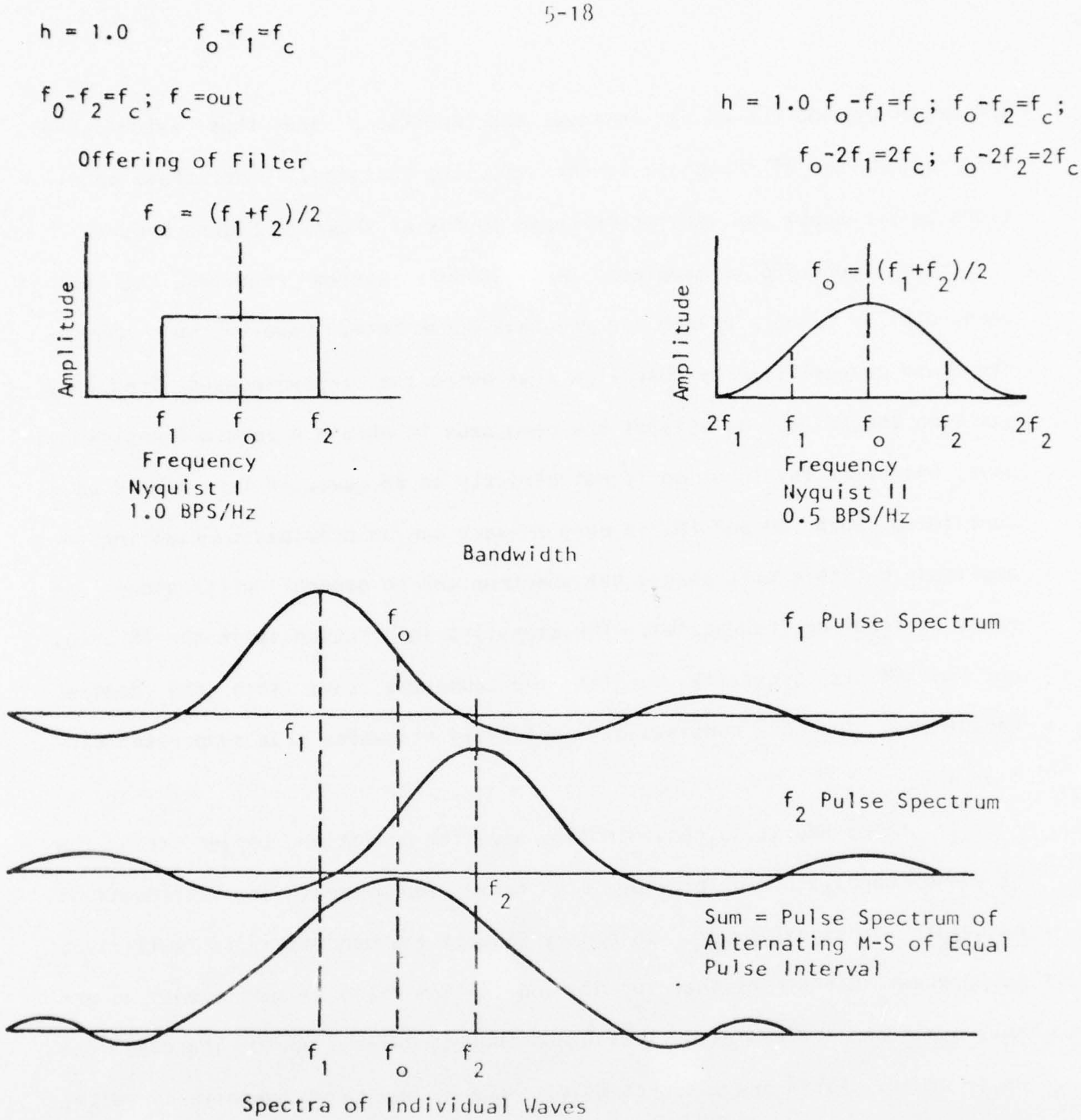


Figure 5.6. Sunde's FSK System

distribution is all important and must be considered, because it determines the error probability. In binary FSK, our ultimate interest is only in whether the frequency deviation from f_o is positive or negative. To evaluate the probability of error, we require the probability that the sign of the frequency deviation is negative when positive was transmitted. The error rate varies from sequence to sequence and within a sequence due to intersymbol interference which produces a variable pedestal for the noise. Thus, it is advantageous to keep the intersymbol interference low, but a compromise with bandwidth available is usually an objective.

The sequence least vulnerable to errors is $(b_n - m) = (b_n + m)$, (alternating marks and spaces) the sequence most vulnerable to errors is $(b_n + m) = -(b_n - m)$, (two marks separated by one space or two spaces separated by one mark).

The situation differs from cases where the detection process is linear. In FSK we cannot make a simple identification of the minimum probability of error with the minimum S/N ratio since there are two ratios to consider; both the incident noise and its derivative enter into the evaluation.

The average probability of error depends on the choice of deviation ratio and method of detection (coherent or non-coherent). Sunde's system wastes considerable power in the f_1 and f_2 frequency components, and on a basis of the same value of average total signal power and pulse rate, is worse than DSB-AM. However, on an equal maximum steady-state power and equal pulse rate, the FSK average total power is twice that of DSB-AM.

The minimum error rate, with conventional FSK, is obtained with a pair of filters at the receiver matched to the two possible signals. It has been shown that optimum coherent conventional FSK is obtained with a deviation ratio h = 0.715 and that this system requires 0.8 dB less S/N ratio than

binary orthogonal signaling (deviation ratio h = 0.6716). In conventional m-ary FSK, the frequencies are usually chosen so that they are mutually orthogonal. This is accomplished by having the frequencies differ by an integer number of cycles over the symbol time. P_e versus S/N curves for the various conventional FSK schemes are shown in Figures 5-7.

5-4. PARTIAL RESPONSE SIGNALLING

Although the initial paragraphs in this section treat partial response signalling as baseband signalling, the subject matter is presented here to introduce the scheme to FSK systems.

5-4.1 DOUBLE DOTTING. With a raised cosine spectrum we have defined the Nyquist II rate as $2f_c$ with the bandwidth also defined as $2f_c$ or in general terms $(f_c + f_x)$. We avoid intersymbol interference if we signal at $2f_c$, but there is no real contradiction if by accepting some intersymbol interference we signal successfully at $2(f_c + f_x)$.

Schemes which seem to surpass the Nyquist rate escape by a loophole in logic by having more signal levels at the receiver than were transmitted, and decoding of the received signal is unique only when certain of the other signal values can be detected.

If we transmit binary pulses (two-levels) at twice the Nyquist II rate and receive 3-levels, translatable into binary, the received signal can be considered to be a ternary signal in terms of S/N ratio at the receiver input. On a maximum steady state power limit basis, for a doubling of the information bit rate, we pay a penalty of 6 dB in S/N; which is an improvement over the 9.5 dB reduction associated with quaternary Nyquist II signaling.

(a) Coherent FSK with optimum deviation (h=0.715) and matched filters

(b) Sunde's h=1.0 coherent, optimum (cosine)+ shaping

(b) Also is Lenkurt duobinary on equal bit-rate basis

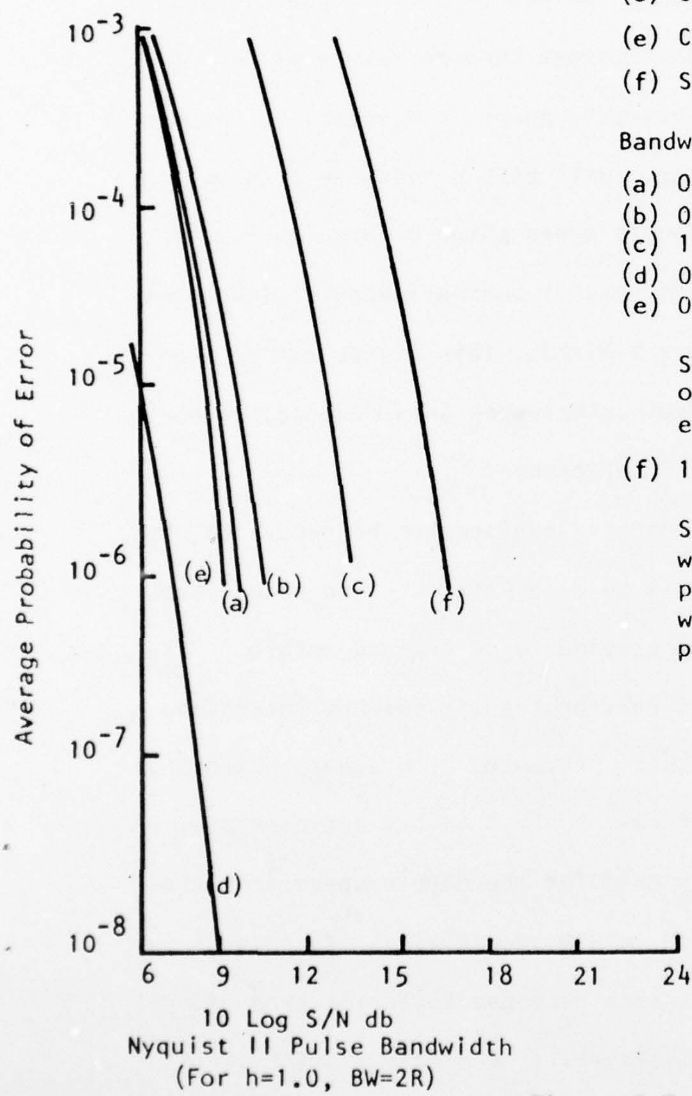
(c) Lenkurt duobinary on equal pulse length basis

(c) Singer duobinary equal bit rate basis

(d) CPBFSK, h = 0.5, coherent, unconstrained MLD

(e) CPBFSK, non-coherent, average matched filters

(f) Singer duobinary,equal pulse length basis



Bandwidth Efficiency/Nyquist II

(a) 0.58 BPS/Hz
(b) 0.5 BPS/Hz; 1.0 BPS/Hz Duobinary
(c) 1.0 BPS/Hz
(d) 0.67 BPS/Hz
(e) 0.67 BPS/Hz

SSB-AM 1.0 BPS/Hz curve is 0.2 db worse than (b) on equal pulse length basis; 2.8 db better on equal BPS

(f) 1.0 BPS/Hz

SSB-AM duobinary 1.0 BPS/Hz curve is 0.2 db worse than curve (f) on max. steady-state power basis, equal pulse length of 0.2 db worse than curve (d) on max. steady-state power basis, equal bit-rate

Figure 5.7. FSK P(e)

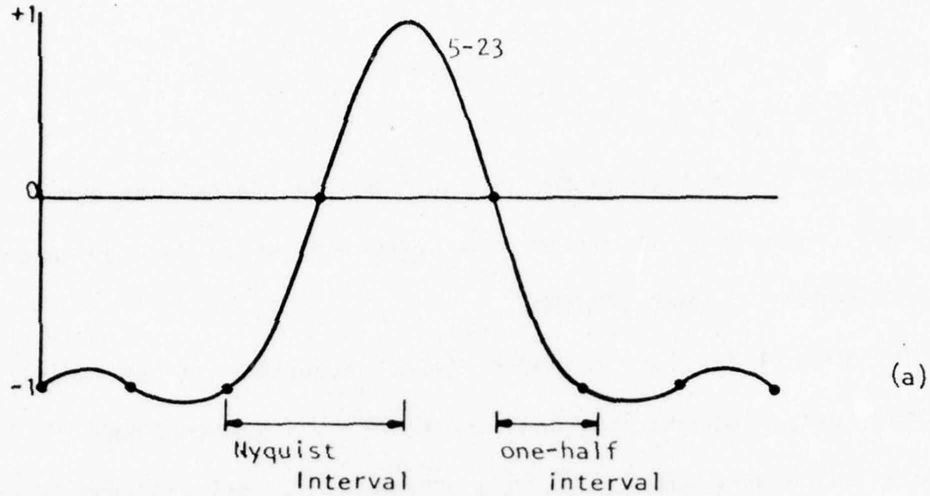
On an equal information bit rate basis the 3-level partial response penalty is only 3 dB when compared with the binary stream.

Originally, in telegraphy days, this technique was called "double-dotting" and limited to transmission at twice the information bit rate. In today's digital systems m-ary Nyquist II signals are also applied on a partial response basis.

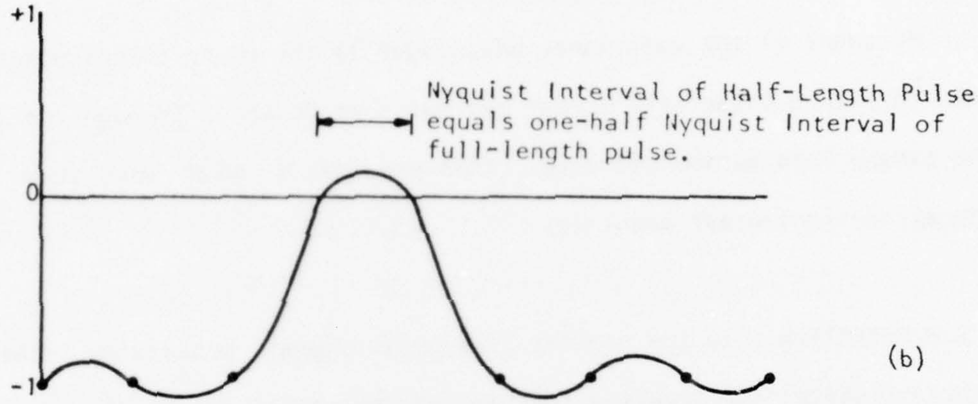
If a rectangular pulse is applied to a raised-cosine channel [Figure 5.8(a)] at the Nyquist II rate, the signal passes through values of zero and 0.5 at instants one-half the Nyquist interval apart. However, a double speed pulse applied to the same channel, will have a response as shown in Figure 5.8(b), and a random series of double speed polar binary pulses applied to the same channel produces a wave that at one-half Nyquist intervals has values of -1, zero, or +1 [see Figure 5.8(c)]. This signal can be considered to deliver 3-level samples without intersymbol interference, or more accurately, with controlled intersymbol interference.

It will be noted that the original binary waveform can be recovered by sampling the wave and interpreting a +1 to be a mark, a -1 to be a space, and the zero value to be opposite of the previously determined state. The pulse response preserves the spacings between transitions but intersymbol interference between the transitions is not prevented. However, sampling takes place at the transition points, at which only 3 levels are possible.

We mentioned above that the penalty paid for the double-speed in terms of equal input binary bit rate, was 3 dB in S/N sensitivity. This value is obtained with the assumption that the partial response filtering is a function of the transmitting channel filter $x(\omega)$. A more efficient partial response configuration, which is only 2.1 dB more sensitive to S/N ratio than binary antipodal, is to provide a matched filter at the receiver by re-



Full-length Pulse Response/Raised Cosine Channel



Half-Length Pulse Response/Raised Cosine Channel at Full-Pulse Length

Random Double-Speed Pulse Response Input

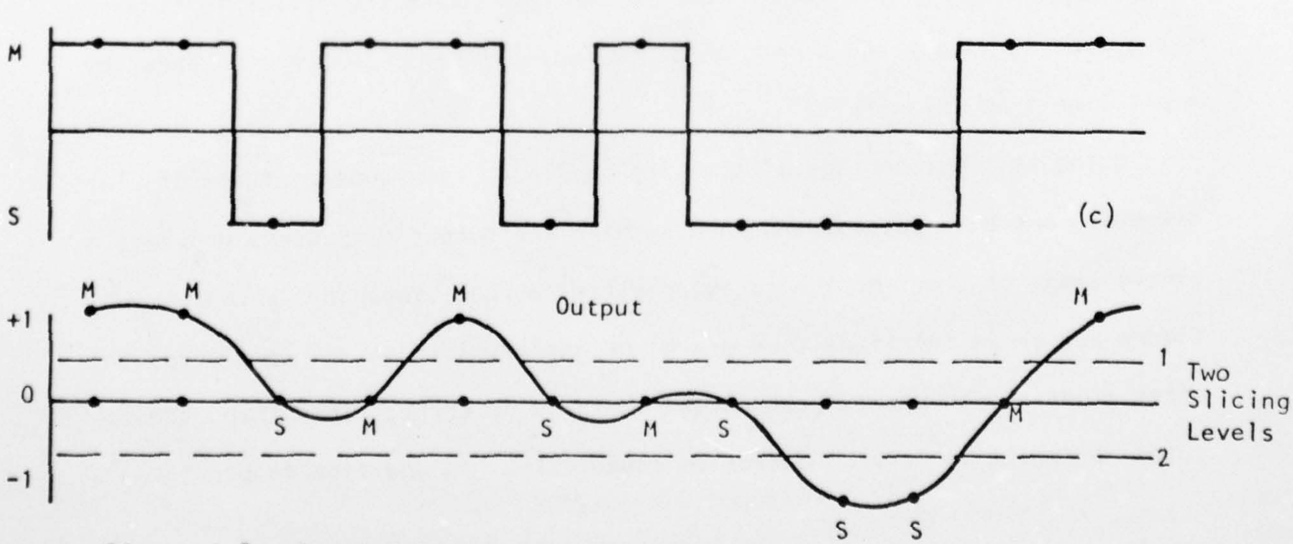


Figure 5.8. Partial Response

placing $x(\omega)$ at the transmitter by $x(\omega)^{\frac{1}{2}}$ at the transmitter and $x(\omega)^{\frac{1}{2}}$ at the receiver. The receiver filter now reduces noise as well as completing the partial response signal shaping.

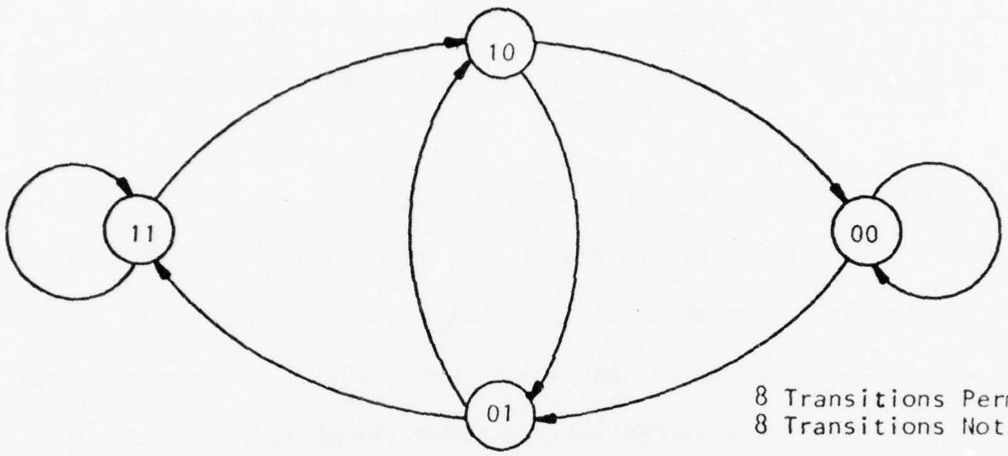
The 3-level partial response (also referred to as double-dotting, pseudo-ternary, pseudo-bipolar, and class IV signaling) has additional advantages which make the 2.1 dB loss a reasonable tradeoff, and in one class of detection, the tradeoff approaches zero.

Foremost of the additional advantages is the power distribution, which due to its bipolar-like format has nulls at DC and a frequency of one-half the binary information bit rate. This provides a good adaptation of the signal to transformer coupling.

5-4.2 DETECTION. In the partial response signal processing scheme, the present state is always a function of the present binary input state plus the previous binary input state.

In Figure 5.8, it is shown that an input pulse one-bit (double-speed) wide, produces an output at two successive bit times, each equal to one-half of the peak output amplitude. An input pulse two-bits wide can be considered as two one-bit pulses in succession, and since the filter is linear, the output is simply the summation of two one-bit responses, one delayed by a bit time from the other.

Using this methodology of treating any input as a superposition of individual one-bit pulses, and then summing the output responses, provides a simple means of plotting the receiver filter output for any given input. Figure 5.9 shows the transition graph. Eight transitions are permitted; the other eight should never occur unless there is an error. This then constitutes a built-in error detection capability. In addition to providing a



(1) A positive peak at one bit time may not be followed by a negative peak at the next bit time.

(2) A negative peak at one bit time may not be followed by a positive peak at the next bit time.

(3) When a positive peak is followed by a negative peak at some later time, they must be separated by an odd number of center samples.

(4) When a negative peak is followed by a positive peak at some later time, they must be separated by an odd number of center samples.

(5) When a positive peak is followed by a positive peak at some later time, they must be separated by an even number of center samples.

(6) When a negative peak is followed by a negative peak at some later time, they must be separated by an even number of center samples.

Figure 5.9. 3-Level Partial Response Transition Graph

means of monitoring system performance, the detection of transition violations can also be used in a control loop to correct for errors in receiver sampling time.

A transition violation detector is not an ideal error detector in that one can postulate pairs of errors which do not violate the transition rules.

5-4.3 DUOBINARY (Refer to Figure 5.10) is a special case of 3-level partial response. The input binary signal is first applied to a differential encoder on the basis that an input space pulse is represented by a "transition" in the output, and an input mark pulse is represented by "no-transition." When the differentially encoded signal is applied to the partial response filter, an input "transition" is represented by the zero value and a "no-transition" by either the peak + or - values.

Thus, after differential and partial response encoding, a binary input space is represented by the zero value, and a binary input mark by either + or - values.

A simple method of detection is by full-wave rectification which folds the - level over the + level. The resultant waveform can then be sliced at the one-half amplitude level to recover the original mark-space binary input intervals without sampling. This method, however, produces an output containing considerable pulse distortion.

This degradation can, of course, be avoided by center bit sampling with two slicers which determine whether the signal at the sampling instant is in the center region (representing a binary input space) or outside the center region (representing a binary input mark).

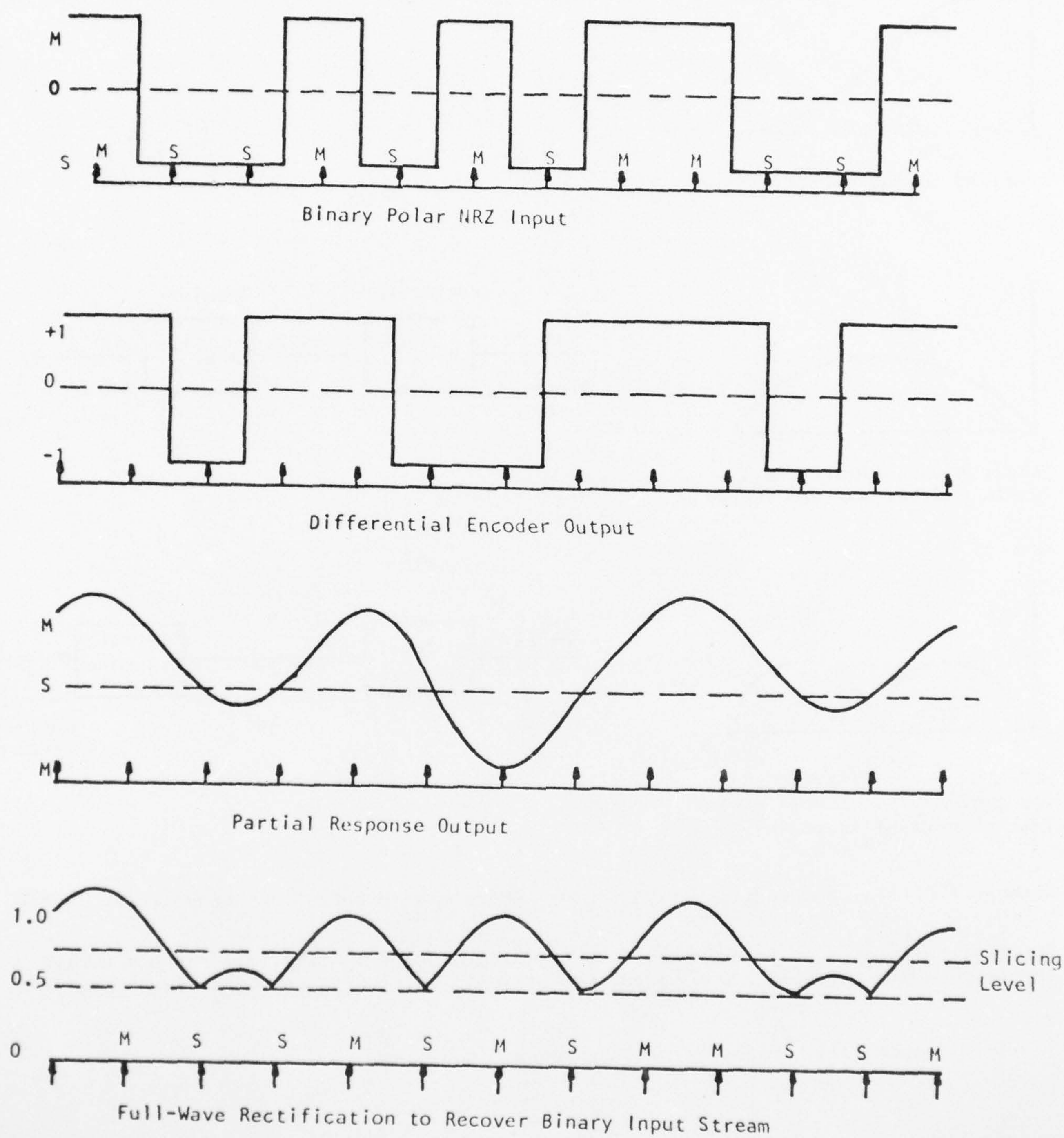
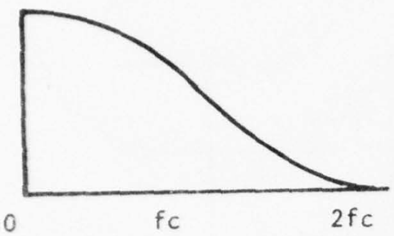
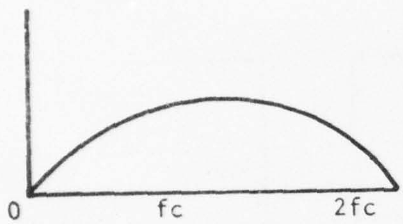


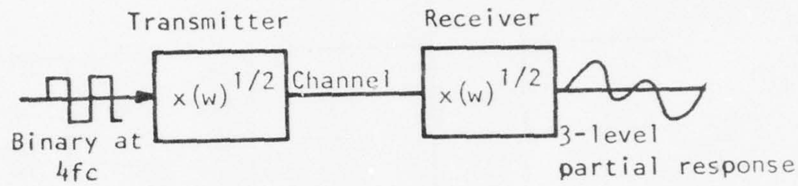
Figure 5.10. Duobinary Signalling

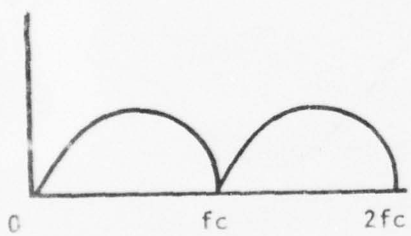


Raised Cosine Filter

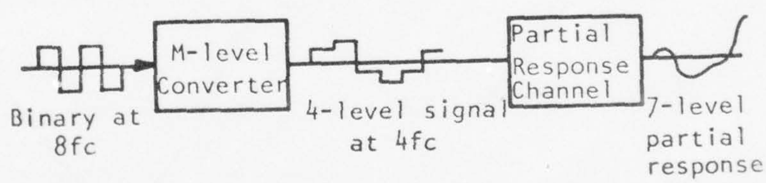


Spectrum when binary input
is 4fc 3-level partial response





Spectrum when binary input at
8fc applied to 4-level encoder
7-level partial response



Number of Partial Response Levels $M_p = 2m-1$ where m = number of baseband signal levels

Spectral Nulls occr at frequencies $f_k = (\frac{N}{K})(4fc)$ where K = 0,1,2,3... and N = Binary information bit multiplication factor, relative to full-pulse response signalling (2fc)

Figure 5.11. M-ary Partial Response Signalling

AD-A065 744 PURDUE UNIV LAFAYETTE IND SCHOOL OF ELECTRICAL ENGI--ETC F/G 17/2
ENGINEERING FUNDAMENTALS FOR BASE WIRE TRANSMISSION.(U)
OCT 78 H K THAPAR, B J LEON, T H WEAVER F30602-75-C-0082
UNCLASSIFIED 1842 EEG-EEIC-TR-79-7 NL

3 OF 3
AD
A065744

END
DATE
FILMED
5-79
DDC

5-4.4 M-ARY PARTIAL RESPONSE SIGNALING (refer to Figure 5-11). m-level baseband signals can also be applied to the partial response channel on the basis of twice the m-level pulse rate. The resultant number of levels in the partial response signal is then

$$m_p = 2m-1 \tag{5-3}$$

where m is the number of levels in the m-level baseband signal input to the partial response channel; or in terms of the information bit rate multiplication factor (compared to the original binary signal), the number of levels in the partial response signal is

$$m_p = 2^{(N/2+1)}-1 \tag{5-4}$$

where N is the binary information rate multiplication factor (full-pulse response signaling).

It can be shown that spectral zeros will occur at frequencies

$$f_k = (\frac{K}{N})\ (4f_c) \tag{5.5}$$

where K = 0, 1, 2, ...; N is the binary information bit rate multiplication factor, and f_c is the Nyquist I frequency of the raised cosine filter.

Thus, application of a 4-level baseband signal to a partial response channel would produce a partial response signal carrying four times the information rate of a binary full-response channel, having seven levels, and having spectral nulls at DC, mid-band, and the top edge of the filter (0, f_c, and $2f_c$ of the raised cosine filter).

5-4.5. PROBABILITY OF ERROR AND BANDWIDTH EFFICIENCY. Figure 5.12 compares 3-level partial response and binary polar (NRZ) signalling performance in

curve (a) Binary polar NRZ full-pulse response

curve (b) 3-level partial response on equal bit rate basis with $x(w)^{1/2}$ split filters

curve (c) 3-level partial response on equal bit rate basis with single x(w) filter

curve (d) 3-level partial response on equal pulse-length basis

NOTE THAT:

(1) 3-level partial response on equal bit rate basis, with split $x(w)^{1/2}$ filters, and maximum likelihood detector, would approach curve (a)

(2) duobinary would have error probability of some quantity less than twice that of 3-level partial response when coherent detection is utilized

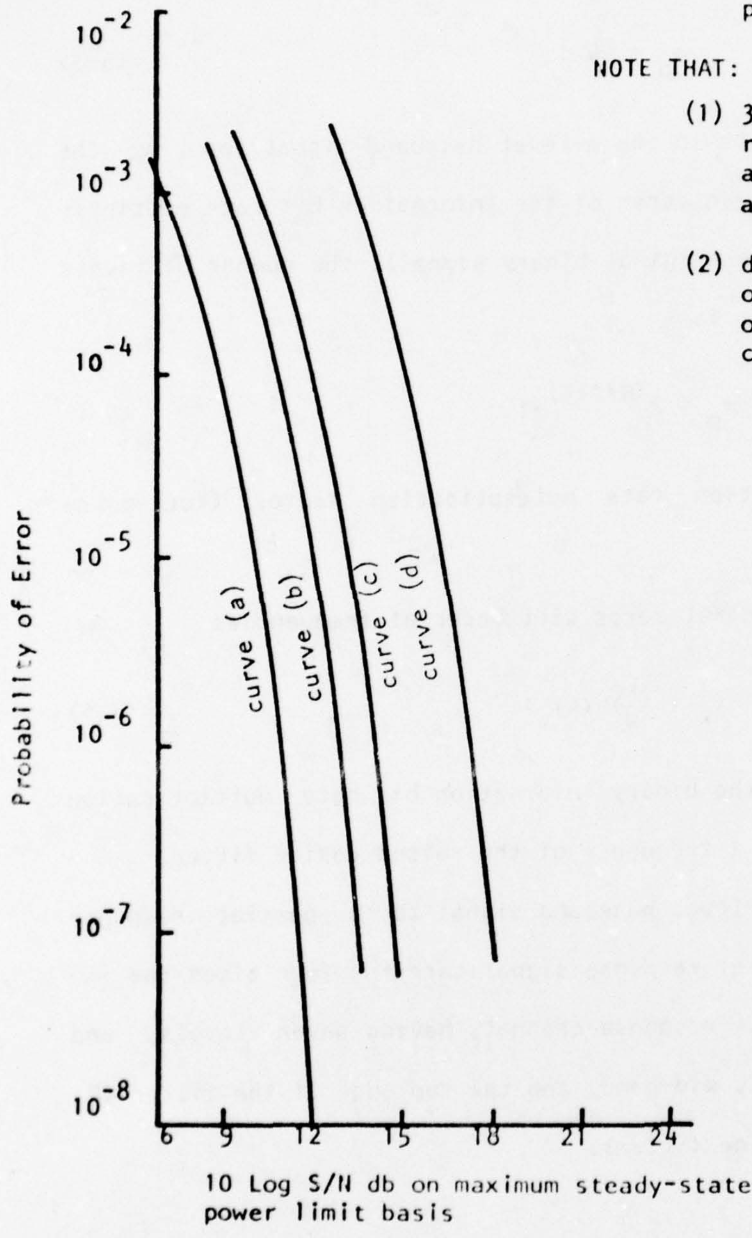


Figure 5.12. Probability of Error 3-Level Partial Response

terms of the probablity of error versus S/N; as can be seen, one does not obtain the increased information transfer rate without cost; and that cost is, of course, a higher bit-error-rate for a given S/N. Figure 5.13 compares m-ary partial response signalling to m-ary polar signalling in terms of bandwidth efficiency versus S/N; as can be seen, at any number of m-levels, partial response is more efficient.

5-4.6 PARTIAL RESPONSE FSK. This form of signalling has been successfully applied to digital signalling over the 4KHz voice channel. The common application is that of a dual-speed modem; i.e., the lower speed uses conventional FSK and the higher speed (2 times lower speed) uses partial response signalling. When the voice channel is quiet (high S/N) the double-speed is used; when the channel is noisy, the single-speed is used.

For example, one system operates at 1200 b/s (low-speed) using a deviation ratio of h = 1.0 (f_1 = 1200 Hz, f_2 = 2400 HZ) meeting the Sunde System requirement for continuous phase at the switching instants. The FSK signal is generated by skewing the frequency of a single oscillator. The resulting signal spectrum is then grouped closely about the average carrier frequency and avoids the power loss in the h = 1.0 discontinuous phase case caused by the mark and space frequency steady-state components. In the 2x rate mode (2.4 Kb/s) the extreme oscillator settings remain the same but the modulating baseband signal is differentially encoded and then partial response encoded (3-level) utilizing the split-filter method, $x(\omega)^{\frac{1}{2}}$ at the transmitter and $x(\omega)^{\frac{1}{2}}$ at the receiver. This duobinary signal now skews the oscillator between 1200 Hz, 1800 Hz and 2400 Hz with the two extreme values representing the binary input mark and the center value representing a space. Thus, "double-dotting" is obtained with FSK in a manner similar to that utilized

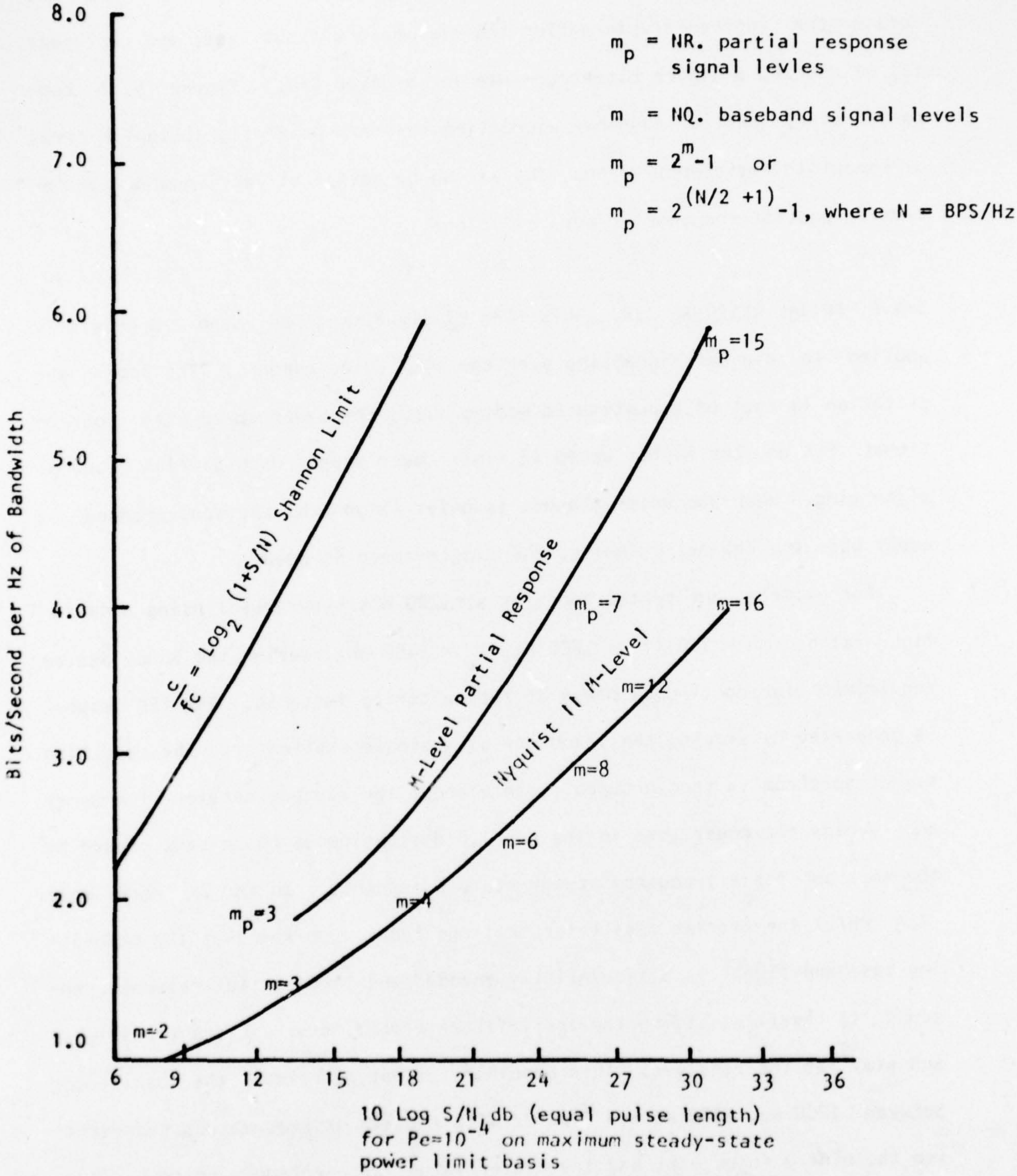


Figure 5.13. Bandwidth Efficiency Partial Response Signals

at baseband. Compared to binary h = 1.0 discontinuous phase FSK, 3-level FSK would suffer a 6 dB penalty in S/N sensitivity, but the continuous phase signal generation provides twice as much effective signal energy and thus the net penalty is 3 dB for the continuous phase duobinary scheme. On an equal bit rate basis, it suffers a 3 dB penalty compared to its single-rate (1.2 Kb/s) continuous phase mode and is equal to a binary discontinuous phase h = 1.0 scheme.

5-5 PHASE-MODULATION

5-5.1 SINUSOIDAL PHASE MODULATION. Phase modulation is a modulation scheme in which the amplitude of the carrier wave remains constant, while the phase is varied in accordance with the modulating signal. The difference between the instantaneous phase of the phase modulated signal and the phase of the unmodulated carrier wave is defined as the instantaneous phase deviation. Under the phase modulation scheme, it is the instantaneous phase deviation that is proportional to the message. The modulated waveform for PM and FM look alike; therefore, the modulation function must be known in order to distinguish between the two.

5-5.2 PHASE SHIFT KEYED SIGNALING. In sinusoidal angle modulation systems, the carrier phase may be expressed as the integral of the instantaneous carrier frequency, or conversely, the instantaneous carrier frequency is the derivative of the carrier phase. If the phase is represented by a sine function, the frequency is represented by a cosine function and vice versa.

For non-sinusoidal angular modulation the difference is more striking. Squarewave continuous phase FSK produces triangular phase modulation, but on

the other hand, except for multiples of 2π radians, squarewave phase modulation results in sharp pulses of frequency modulation.

When data is transmitted by m-ary phase modulation, all phase steps are restricted to within a range of $\pm 180^{\circ}$. Such a wave can be considered to be two amplitude modulated waves with their carrier phases in quadrature. Signals modulating the carriers are m-level squarewaves with amplitudes proportional to the sines and cosines of the phase values. The spectra of such phase modulated waves can thus be considered as the superposition of two amplitude modulated spectra (each of which is a DSBSC wave).

As with FSK, there are two cases of phase shift keying (PSK); discontinuous or switched-phase and continuous phase. If the phase changes can be represented by a staircase function; i.e., instead of returning to a previous smaller value, the phase continues to advance in discrete steps, and this phase function can be resolved into a sum of a straight line and a periodic component, the straight line indicates a shift in carrier frequency, and the wave can be considered to be continuous.

Phase modulated systems, of course, must employ coherent detection techniques, which may take either of the two general forms, i.e., transmission of a reference wave or recovery of the carrier from the transmitted data wave. In the latter case, there is the uncertainty of whether or not the recovered reference is 180° out of phase with the incoming signal (2-level PSK) or in the case of m-ary PSK of the demodulator arbitrarily choosing any one of the m-phases as a reference.

In 2-level PSK, choice of the 180° out of phase condition as a reference would provide a complementary data stream at the detector output. In m-ary PSK the resultant detector output depends on which of the m-phases is chosen as the reference and which of the m-level modulation schemes is util-

ized.

However, regardless of modulation scheme, most PSK modem designers have concluded that differentially coherent PSK is the best compromise for most applications.

In 2-level differentially coded PSK, one input data state may be represented by a change in phase of 180° from that of the previous bit time interval and the opposite data state by no change in phase; or one input data state may be represented by a change in phase of 90° in the positive direction from that of the previous bit time interval and the opposite data state represented by a change in phase of 90° in the negative direction from the previous bit time interval.

In 4-level differentially encoded PSK the signal diagram is usually one of two constellations depending on the choice of modulation scheme.

In one of the schemes, usually referred to as 4-level switched PSK, the incoming binary data stream is first encoded (Gray code) into a 4-level tuplet pulse stream in which each of the tuplet then rotates the phase vector from the previous pulse interval by $k\pi/2$ radians, where k is

Tuplet (Gray Coded)	k	
00	0	(no change)
01	+1	($+90^{\circ}$ change)
11	+2	(180° change)
10	-1	(-90° change)

In the other scheme, usually referred to as QPSK, the incoming binary data stream is first split into two separate pulse streams on an every-other-bit basis, with each of the separate pulse streams having a pulse time

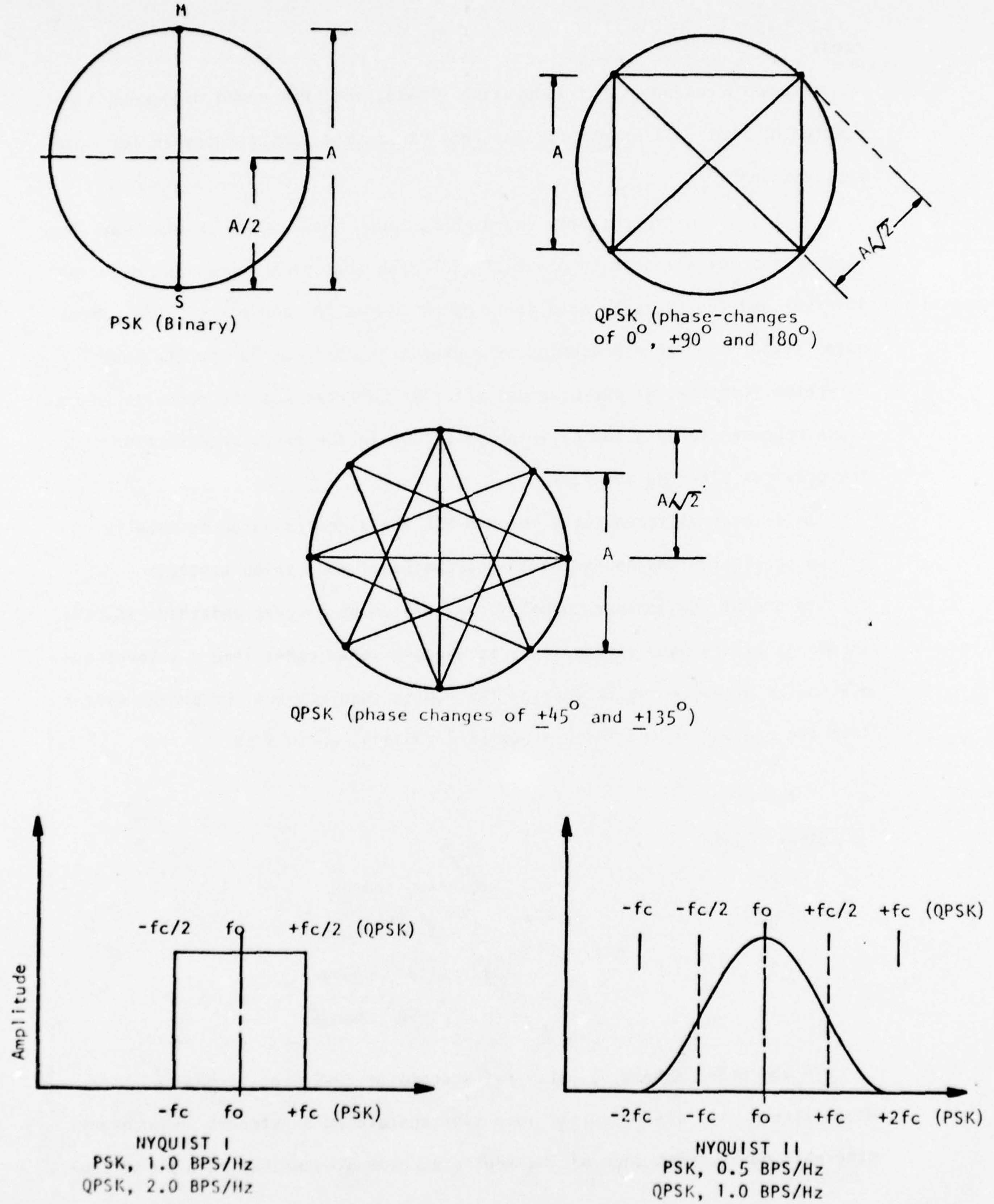


Figure 5.14. Phase Modulation Signal Space Diagrams and Required Pulse Bandwidths

interval of twice that of the original binary data stream. Each of the pulse streams modulates a quadrature carrier on a 2-level differentially encoded basis and the quadrature signals are then summed.

The resulting signal is usually arranged to rotate the phase vector from the previous time interval $\pm k\pi/4$ radian where k is

Tuplet Represented By Summed Signal	k
11	+1 (+45^{o} change)
01	-3 (-135^{o} change)
00	-3 (+135^{o} change)
10	-1 (-45^{o} change)

It has been shown that in a system which compares the phase of adjacent symbols, that the carrier frequency should exceed twice the maximum modulating frequency rather than just greater than the modulating frequency as is the case in non-differential phase detection.

5-5.3 PSK PERFORMANCE. On the basis of a Nyquist I pulse bandwidth, the PSK signal being a double sidebanded signal would require a bandwidth equal to the pulse rate; on a Nyquist II basis, a bandwidth of twice the pulse rate is required. Thus for PSK (2-level) we consider the bandwidth efficiency to be 0.5 BPS/Hz (Nyquist II) and for QPSK, 1 BPS/Hz (refer to Figure 5.14).

In terms of probability of error (Figures 5.15 and 5.16) in the presence of white Gaussian noise, PSK is antipodal and the minimum separation of A (Figure 5.14) requires a pulse amplitude of A/2; QPSK requires a pulse amplitude of $A/\sqrt{2}$, 3 dB greater than PSK. On an equal bit rate basis, QPSK becomes equal in performance to PSK in that QPSK requires only one-half the

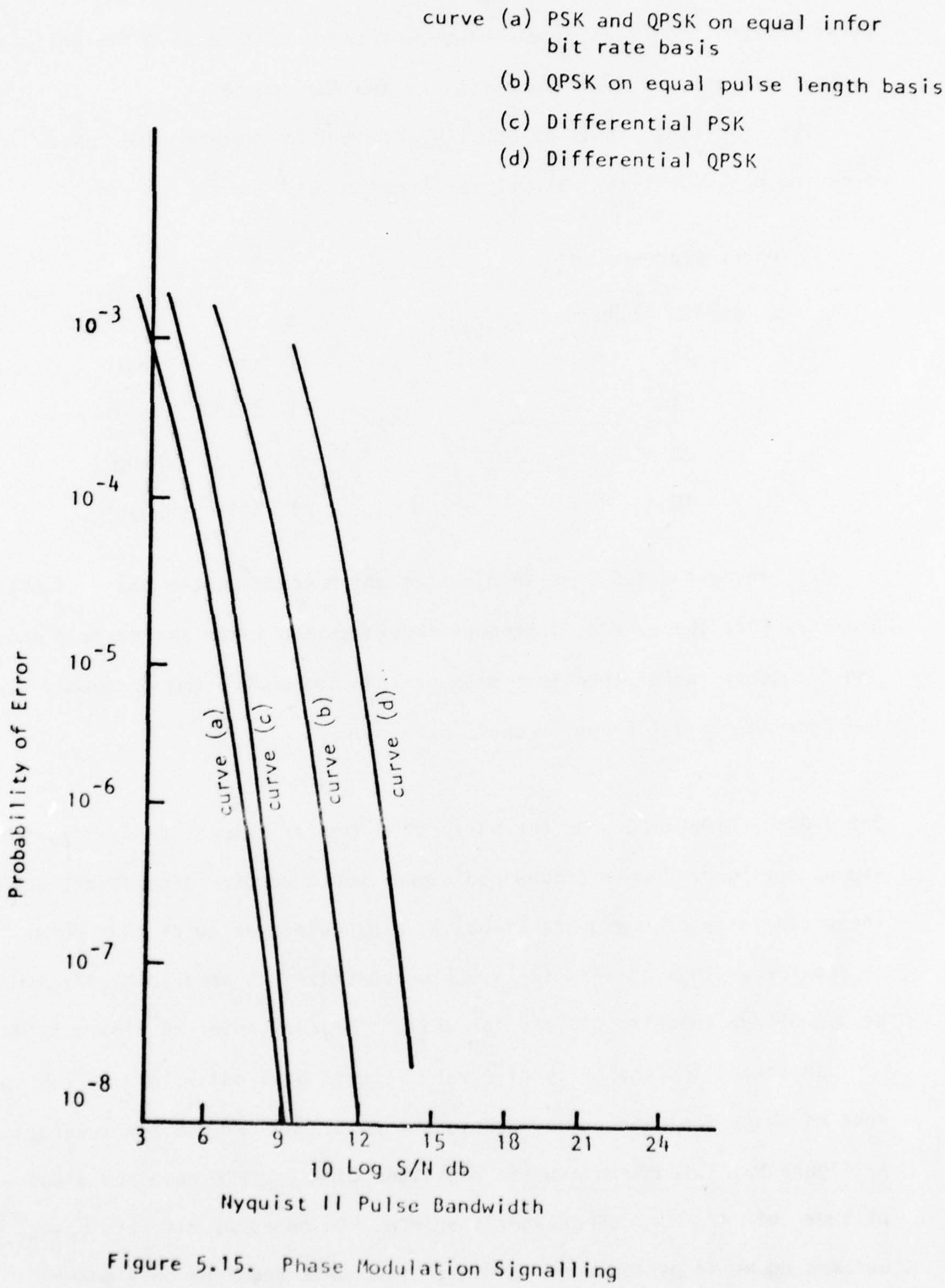


Figure 5.15. Phase Modulation Signalling

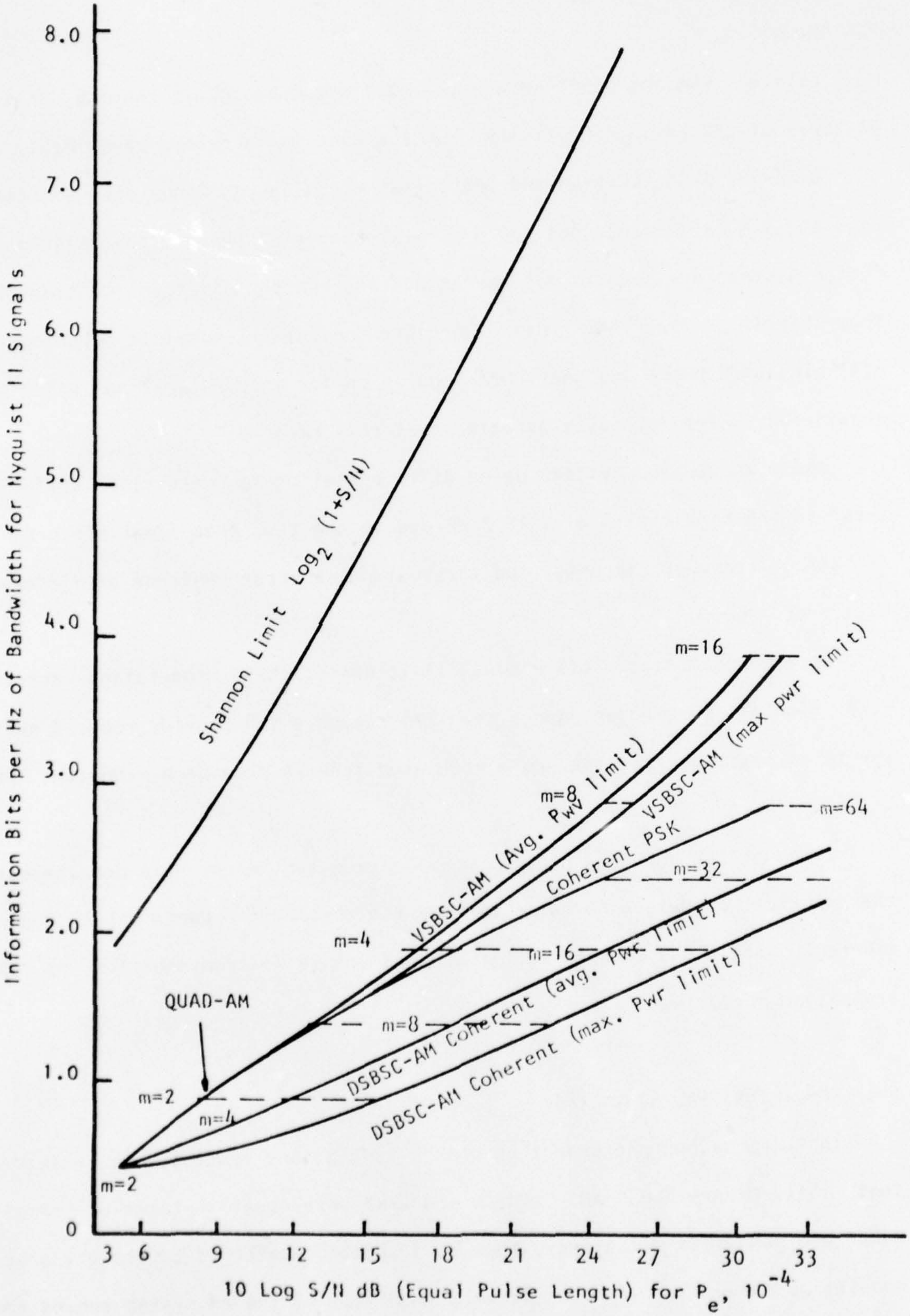


Figure 5.16. Phase Modulation Bandwidth Efficiency

PSK bandwidth.

Coherent PSK and DSBSC-AM are the most advantageous signaling schemes in terms of S/N ratio sensitivity, but the cost is increased bandwidth.

Binary VSB-AM, QUAD-AM and QPSK provide equivalent bandwidth efficiencies and S/N performance but for a greater number of phases, the phase modulation systems are seen to fall along a line which diverges considerably from VSB-AM and QUAD-AM. This divergence indicates a possible advantage of utilizing both phase and amplitude modulation for m-ary signaling which is considered in the following paragraphs of this report.

Phase modulation systems using differential phase detection suffer a loss in S/N sensitivity of 1 to 3 dB due to the less than ideal performance of carrier recovery circuitry and the dual-pulse-error syndrome of differential encoding.

On a maximum steady state power limit basis, phase modulation schemes (and FSK) are more advantageous than AM schemes which provide reduced average total power whereas the angle modulated schemes provide a constant envelope.

Although PSK and QPSK systems provide excellent error rate performance, the practical bandwidth normally exceeds the theoretical bandwidth by a considerable margin due to both amplitude and phase discontinuities at the transition instants.

5-6. TWO-DIMENSIONAL SIGNALING

In the previous paragraphs on phase modulation systems, we observed that while binary VSBSC-AM, QUAD-AM and QPSK were equal in terms of performance and bandwidth efficiency, that at a greater number of levels the phase modulation curve diverges considerably from that of the AM system curves and

that this suggested the possible advantage of using both types of modulation in combination.

Bell and others devoted considerable resources to the study of this possibility resulting in the development and production of several wireline modems. These studies answered the following questions:

a) Given an n-level and the facility to use ℓ-amplitudes and m-phases such that $n = \ell m$, what is the value of ℓ and consequently m, such that the error rate is smallest for a fixed S/N ratio when signaling over a bandlimited additive gaussian channel?

b) How does this optimum $n = \ell m$ compare with phase modulation or amplitude modulation at a fixed S/N ratio?

c) How effective is this modulation scheme in terms of BPS/Hz at a fixed error rate?

d) How close can actual hardware approach theoretically predicted results?

The studies were further limited to signal constellations in which the total amount of information carried by the signal is given by

$$I/T\ (\mathrm{Log}_2 1 + \mathrm{Log}_2 m) = I/T\ (\mathrm{Log}_2 n) \tag{5-6}$$

in which the amplitude symbols are represented as b_n and the phase symbols as θ_n. The receiver processes the sum signal $s(t)+\omega(t)$ to obtain estimates of b_n and θ_n and reception is accomplished in two parallel detectors denoted as E, the envelope detector and Φ, the phase detector.

The E detector operates by first computing the envelope of $s(t)+\omega(t)$ and synchronously sampling the result every T seconds. The Φ detector com-

putes the phase of the signal (plus noise) which is also synchronously sampled to derive estimates of θ_n. In all operations it was assumed that correct timing information was available to the receiver and as long as s(t) has a non-vanishing envelope, the measurements are unambiguous. Symbol detection was obtained by comparison with fixed thresholds.

Results of the study question a) are as follows with "yes" as the overall answer to the question. Is there an optimum ℓ and m for a given n where $n = \ell m$?

Table 5-1 OPTIMUM ℓ AND m FOR VALUES OF n

	Optimum	Optimum	dB	dB	dB	dB	dB
n	ℓ	m	DEΦd	DEΦc	DΦd	DΦc	Da
4	2	2	7.0	7.0	5.3	3.0	7.0
8	2	4	9.6	8.3	11.2	8.3	13.2
16	2	8	13.5	11.5	17.3	14.2	19.3
32	4	8	16.9	15.6	23.2	20.2	25.3
64	4	16	20.4	18.5	29.2	26.2	31.3
128	8	16	23.5	22.2	35.2	32.3	37.4

where

n is the total number of levels, ℓ is the number of amplitude levels in a combined system,

m is the number of phases in a combined system,

D is the minimum value of envelope in the absence of noise in terms of total average power,

EΦd indicates a combination system with differentially coherent detection,

$E\phi c$ is a combination system using coherent detection,

ϕd indicates a phase modulation system with differentially coherent detector,

ϕc is a phase modulation system with coherent detection, and

a indicates an amplitude modulated system (DSB).

The answers to question b) are also contained in the foregoing tabulation. For $n = 4$, all phase modulation is better than the optimum combination by 1.7 dB so there is no advantage in splitting levels. For $n = 8$, the combination $\ell = 2$ and $m = 4$ is preferable by 1.6 dB over 8 phases and by 3.6 dB over 8 amplitudes.

From the data given it appears that it is always advantageous to use the optimum ℓm combination when $n \geq 8$ and is more pronounced as n becomes large.

Question c) concerned the bandwidth efficiency in terms of BPS/Hz.

$$R = \text{BPS/Hz} = \frac{\text{Log}_2 n}{T\beta} = \frac{P}{\beta} \log_2 n \tag{5-7}$$

where

P is the symbol signaling rate and β is the required bandwidth.

For ideal double sideband systems $\frac{P}{\beta} = 1.0$ and for ideal single sideband systems $\frac{P}{\beta} = 2.0$, thus

$$R_{DSB} = \text{Log}_2 n \quad \text{BPS/Hz} \tag{5-8}$$

$$R_{SSB} = 2\ \text{Log}_2 n \quad \text{BPS/Hz} \tag{5-9}$$

and <u>for</u> <u>SSB</u>

$$S/N \text{ dB} \simeq 20\mathrm{Log}_{10} n = \frac{20}{\mathrm{Log}_2 10} \mathrm{Log}_2 n \tag{5-10}$$

and since

$$R = 2\mathrm{Log}_2 n \text{ we have that } R \simeq \frac{\mathrm{Log}_2 10}{10} (S/N) \text{ dB} \tag{5-11}$$

<u>for</u> <u>DSB</u>

$$S/N \text{ dB} \simeq 20\mathrm{Log}_{10} n = \frac{20}{\mathrm{Log}_2 10} \mathrm{Log}_2 n \tag{5-12}$$

and since $R = \mathrm{Log}_2 n$

we have

$$R \simeq \frac{\mathrm{Log}_2 10}{20} (S/N) \text{ dB} \tag{5-13}$$

<u>for</u> <u>combined</u> <u>AM/PM</u>

$$S/N \text{ dB} \simeq 10\mathrm{Log}_{10} n = \frac{10}{\mathrm{Log}_2 10} \mathrm{Log}_2 n \tag{5-14}$$

and since $R = \mathrm{Log}_2 n$

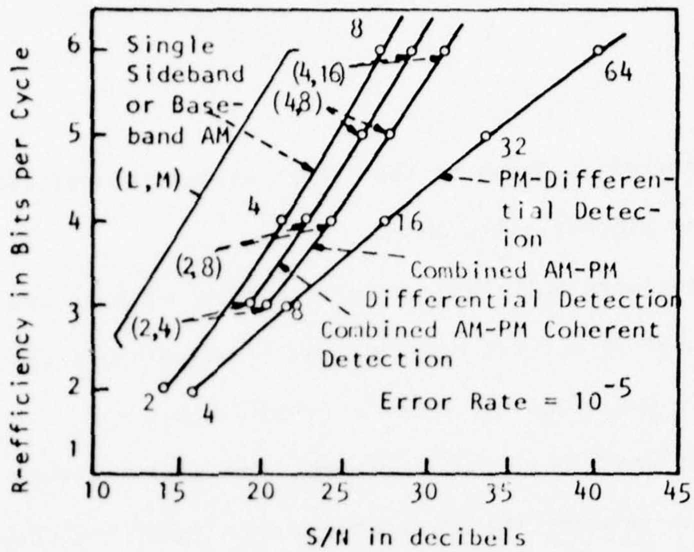
we have

$$R \simeq \frac{\mathrm{Log}_2 10}{10} (S/N) \text{ dB} \tag{5-15}$$

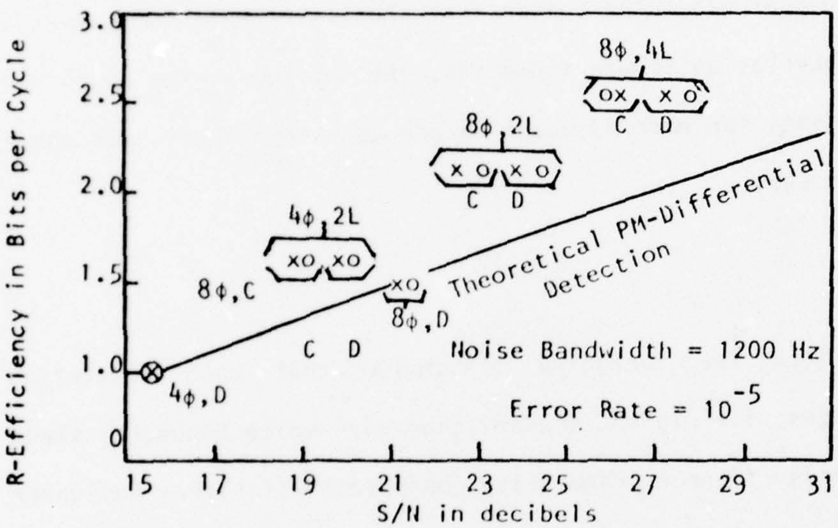
and finally for <u>Shannon's</u> <u>Limit</u>

$$R_{capacity} \simeq \mathrm{Log}_2 S/N = 1/3(S/N) \text{ dB} \tag{5-16}$$

Resulting bandwidth efficiency curves and error performance curves are shown in Figure 5.17a.

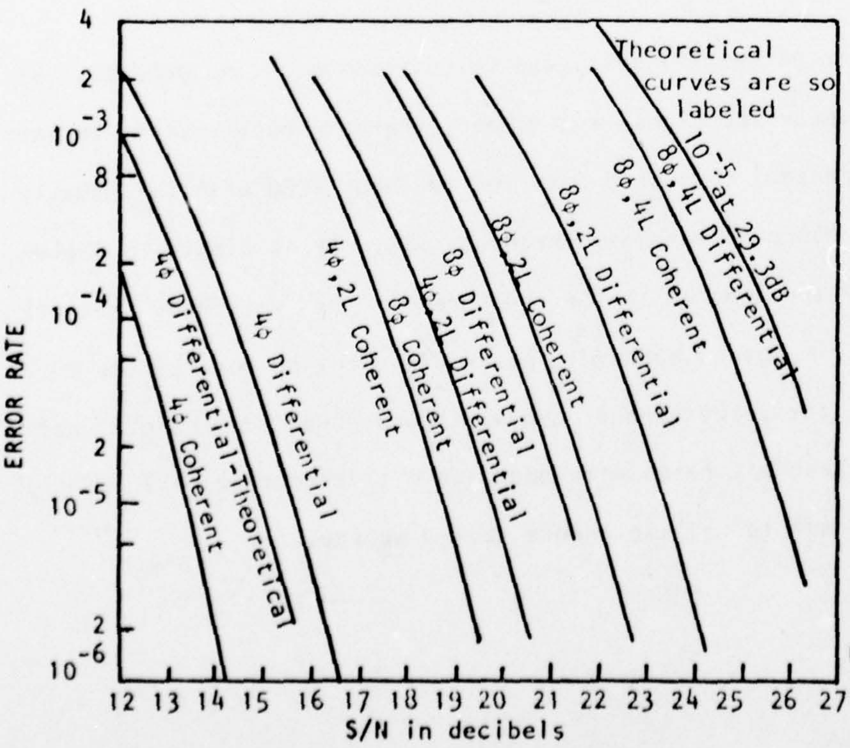


(a) Theoretical Efficiency vs. S/N for Various Systems



(b) Performance Comparisons

x = Actual
o = Theoretical



(c) Performance Curves Actual and Theoretical

Figure 5.17. 2-dimensional Performance

Question d) is answered as shown by the actual versus theoretical performance curves shown in Figures 5.17b and c.

Many wireline modems were built as a result of these studies and eventually two types of signal structures evolved; one in which there was always an equal number of phase positions for each amplitude level, and the other in which the number of phase positions increased with each increase in amplitude in order that the probability of error would be equal for each position based on the higher S/N ratio vailable at the higher amplitude levels. In this type of constellation it was found that the optimum system contained six more phase positions for each adjacent larger amplitude level with three phases in the lowest level.

5-7. SUMMARY

It is apparent from the foregoing discussion that each modulation scheme has advantages for digital transmission over voice channels; these were presented in terms of error probability, bandwidth efficiency and power efficiency.

However, another parameter, cost, was not discussed. In general, of course, the more complex the signalling scheme, the more costly it is to implement. For this reason, low-speed data (up to about 2000 b/s) is usually accomplished using non-coherent (asynchronous) FSK. It is simple to implement and performs better than AM in the presence of the telephone network impulse noise. It is not as bandwidth power efficient as some of the more complex schemes, but the voice channel bandwidth and power constraints permit signalling at these bit rates with adequate bit error rate performance; thus, there is no reason to utilize a more costly scheme.

At 2400 b/s voice channel modems utilize synchronous signalling with a binary AM, FSK or PSK modulation scheme. Again, more complex schemes are not required to transfer 2400 b/s over the voice channel with adequate performance.

At 4800 b/s m-ary (usually 4-level) AM and PSK (QUAD-AM and QPSK) are utilized in order to obtain the required bandwidth and power efficiencies; at 9600 b/s 2-dimensional signalling schemes are required.

The modem complexity and cost are proportional to the bit rates;

a) Low-speed to 1200 b/s $\simeq$ \$500

b) Medium speed to 2400 b/s $\simeq$ \$2000

c) Hi-speed to 4800 b/s $\simeq$ \$5000

d) Hi speed to 9600 b/s $\simeq$ \$8000

SECTION VI

MULTIPLEXING AND CONCENTRATION TECHNIQUES

6-1. GENERAL.

A wire transmission line is capable of transmitting frequencies ranging from dc to approximately 150 KHz without appreciable distortion. They can be used for reliable digital transmission up to over one megabit per second. Coaxial cables and microwave radio systems have higher capacity than wire pairs. Circular waveguides and glass fibers have higher capacity than coaxial cables and microwave radios. Since the voice-frequency range can be considered to be from 300 Hz to 3,300 Hz, only a very small portion of the potential capacity of the various transmission media is utilized when a single voice conversation is transmitted over such facilities. Based on purely economic considerations, it is desirable to make the maximum use of the capacity of the transmission medium. The technique whereby several channels are carried over one transmission path is referred to as multiplexing. In a multiplex system, two or more signals may be combined and transmitted as one signal; at the receiver, the signal is separated (demultiplexed). The signals to be multiplexed may be voice, telegraph, data, video, facsimile, etc. Through the use of multiplexing, Bell System is able to place 12 voice channels (with nominal bandwidth of 4 KHz each) over an open-wire line, 600 to 1860 voice channels on a coaxial cable and microwave system, and several hundred thousand channels (100,000 to 200,000) on a circular waveguide. Similarly, messages with frequency (bit rate) content much smaller than the nominal voice channel frequency (bit rate) can be combined together and transmitted over a single voice frequency channel. A multiplexer, abbreviated MUX, is the device that performs the multiplexing operation.

The term "multiplexing" is also used in connection with systems other than those dealing with voice transmission. For example in a computer sys-

tem, a single channel, called the "multiplexing channel", may be used to transmit/receive the control information for controlling the operation of the various devices, such as the card-reader, line printer, etc. "Multiplexing" is also used in connection with time-sharing systems where a computer is used as an operator (somewhat akin to a human telephone operator) to control the interconnections between several terminals, or console operators, and the computer.

In the following paragraphs, the word "multiplexing" will be used as commonly referred to in the telecommunications jargon, namely, the utilization of a channel by dividing it into two or more channels. There are many forms of multiplexing; among them space division multiplexing, frequency division multiplexing (FDM), and time division multiplexing (TDM) are most commonly used.

6-2. SPACE DIVISION MULTIPLEXING.

Space division multiplexing is a term used to mean the physical packing of many separate paths (channels) into a cable. This scheme is very commonly used in transmission systems. In telephone transmission, several hundred or thousand twisted pairs of wires are bundled into a cable. Each cable therefore carries many phone conversations. In areas with heavy traffic, large channel cross-sections are required. By using space division multiplexing a saving in cost can be obtained by "grouping" together several individual channels into a large channel cross-section. Space division multiplexing does, however, impose restrictions on the layout of the transmission paths since the application of the technique depends on the ability to combine transmission paths into specific routes. Also, the close proximity of channels as a result of packing can lead to interference. Space divi-

sion multiplexing is most often used in conjunction with other multiplexing systems, such as frequency division multiplexing and time division multiplexing. Unlike FDM and TDM schemes, where the message undergoes several signal processing steps, space division multiplexing does not involve any signal processing.

6-3. FREQUENCY DIVISION MULTIPLEXING.

The technique whereby several baseband signals are combined together in the frequency domain and transmitted over a common transmission path is referred to as frequency division multiplexing. Consider, for example, the transmission of a telegraph message and a telephone conversation. The telegraph messages typically have a spectrum extending from dc to 100 Hz, while a telephone subscriber's voice spectrum extends from 200 to 3300 Hz. Thus, the two signals can be combined and transmitted over the 4 KHz voice channel. Even when these two signals are multiplexed, only a fraction of the transmission medium capacity is utilized. Moreover, most of the baseband signals encountered in practice cannot be multiplexed in the same manner as the example of the telegraph message and the voice message; the frequency spectra of these signals overlap. For example, the frequency spectrum of an average telephone caller is the same as that of any other. In order to multiplex several callers over a common transmission path, some sort of frequency translation of the message frequencies is evident. That is, instead of allocating 0 to 4 KHz frequency spectrum for a voice channel, and thus assigning separate transmission paths for each talker, the following scheme can be used: talker A's natural frequency spectrum of 200 to 3300 Hz can be raised to a frequency band extending from 60,200 to 63,300 Hz (60 to 64 KHz voice channel); talker B can be assigned frequencies between 64

and 68 KHz; talker C can be assigned frequencies between 68 and 72 KHz; and so on. This is precisely what is done in frequency division multiplexing. Signals that occupy the same frequency spectrum are modulated with different carrier frequencies so that their frequency spectrums are translated to new non-overlapping frequency ranges. At the receiver, the multiplexed signal can be filtered and demodulated to recover the message. The use of frequency division multiplexing is not just limited to telephone transmission; there are many other applications. For example, FDM is used in radio broadcast. Different frequencies are allocated among broadcast stations such that each station can broadcast its program independently over a common medium. The signal input to the radio set therefore contains a multiplicity of transmitted signals from many stations. At the receiving antenna there is just one composite signal. By using frequency selective circuits in the receiver a particular broadcast station can be selected from the multiplexed signal. Frequency division multiplexing is also used for transmitting narrrowband signals over voiceband facilities. The frequency allocation of microwave radio is also based on a frequency division multiplexing plan. In the United States this frequency allocation is done by the Federal Communications Commission (FCC). Frequency bands are allocated for radio broadcast, television broadcast, common carrier microwave transmission, etc.

6-3.1. ELEMENTS OF A FREQUENCY DIVISION MULTIPLEX SYSTEM. Figure 6-1 is a block diagram of a basic telephone frequency division multiplex system. This can be said to consist of three major components: the West terminal, the transmission line, and the East terminal. Each terminal is made up of several channels, each of which provides for a single two-way conversation

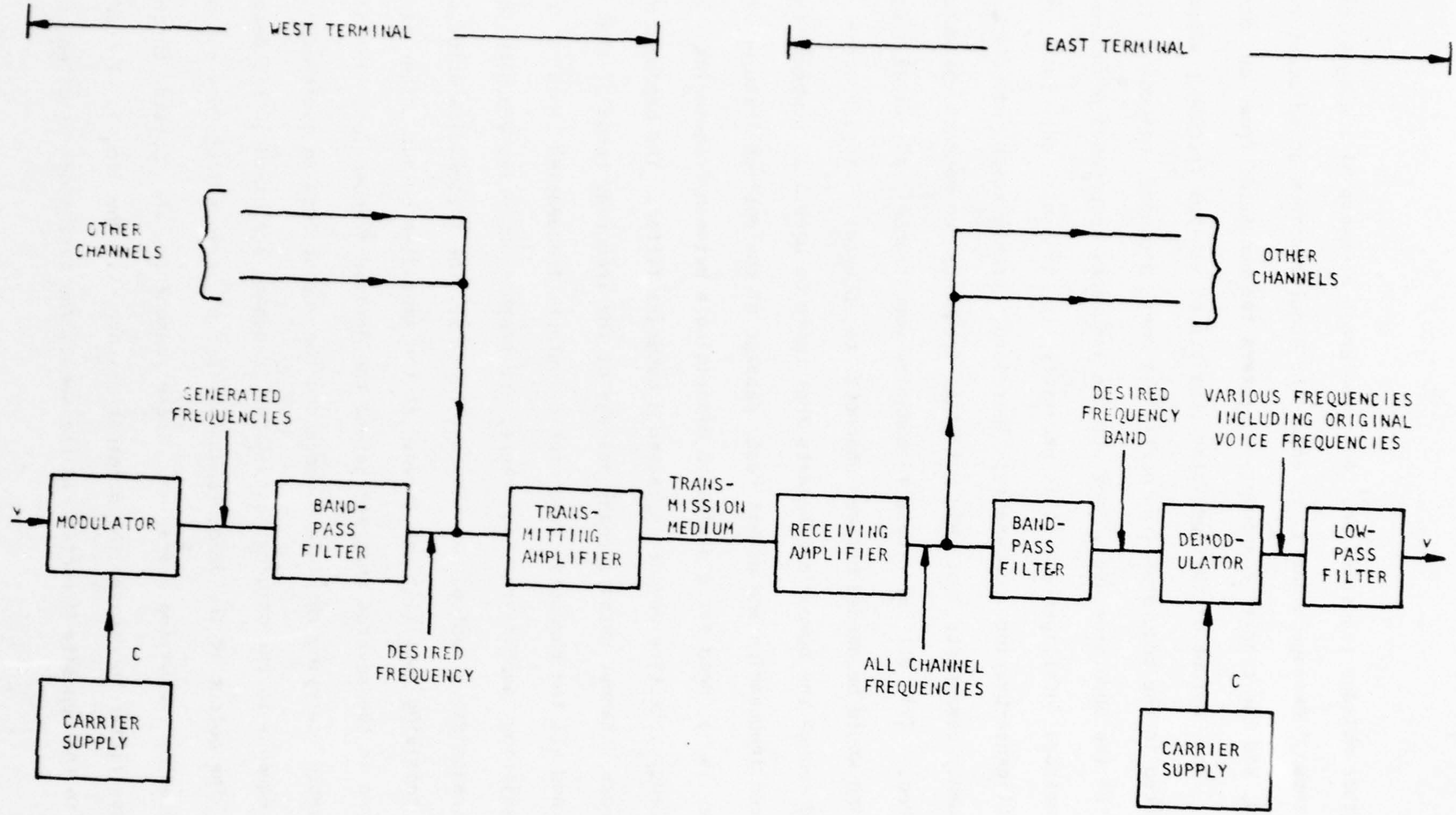


Figure 6-1. Frequency Division Multiplex System

or other message transmission. Only the basic elements of a single channel are shown. Message signals, V, enter the modulator from an adjacent switchboard, and the carrier component, C, enters the modulator from an oscillator. The output of the modulator consists of various frequency components produced in the modulator, the particular ones produced depending on the type of the modulator used. Not all the frequency components produced carry the message intelligence. It is necessary to transmit only one of the intelligence-bearing components; therefore, transmission of only certain frequency components out of all the frequency components produced is desired. (If all of these frequencies were transmitted, large frequency spacing would be needed between channels to prevent interference-- which would reduce the number of channels that could be used.) To suppress the undesired frequencies and permit ready passage of the desired frequency range, there is a need for a device to discriminate between frequencies. Such a frequency selective device is called a bandpass filter. The bandpass filter in each channel permits ready passage of the frequency range of that channel, and all the channels feed their output frequencies into a common transmitting amplifier. From there, all these frequencies are sent out on the transmission medium. At the East terminal, the frequencies enter a common receiving amplifier, as shown. At the amplifier output, other bandpass filters in the several channels select the desired frequencies. The desired frequency carrying the intelligence and the output from an oscillator equal in frequency to the original oscillator frequency are mixed in the demodulator. The output of the demodulator consists of various frequency components produced by the mixing operation. These components are passed through a lowpass filter to recover the original message, V. The process of transmission in the opposite direction is the same. The term modem is often used to

indicate the modulation and demodulation device in the multiplexing system.

There are several methods available for performing the modulation steps at the transmitter. The most commonly used scheme on the commercial carriers is amplitude modulation using a product-type modulator. On long-haul systems, where transmission line costs are of concern, single sideband is most commonly used because of its bandwidth conserving properties. In short-haul systems, on the other hand, the cost of multiplexing terminals is of significance. Due to the simplicity of demodulation inherent in double sideband transmission, the commercial short-haul systems use both sidebands. Thus the nominal voice channel spacing in short-haul systems is 8 KHz instead of 4 KHZ. Also, in the commercial short-haul systems (the N-type carrier), different frequency bands--sometimes referred to as the "high group" and the "low group" bands--are used for the multiplexed signal in each direction of transmission. This interchange of frequency bands at the repeater is called <u>frequency frogging</u>. Frequency frogging reduces the crosstalk by blocking the circulating crosstalk path at each repeater.

So far our discussion of FDM has been concerned with the multiplexing of analog signals. Digital signals, in the form of pulses, can also be frequency-division multiplexed by appropriate pre-processing of the given signal. Digital modulation techniques, such as ASK, FSK, PSK, or DPSK, are used to convert the pulse train into an analog signal and then multiplexed, using FDM, and transmitted over the voice channel or wideband channels, which are themselves derived from multiplexing several voice channels.

On most of the commercial carrier systems, the carrier is not transmitted along with sideband(s). An advantage accrued from suppressing the carrier is the reduction of intermodulation noise and the reduction of transmitter power. With the carrier absent, the intermodulation due to the

sidebands of multiplexed talkers is unintelligible crosstalk; however, when the carrier is present, the intermodulation due to the carrier and a sideband can produce intelligible crosstalk, and be disturbing to the subscribers. There is, however, a price paid for this reduced intermodulation due to suppressed carrier transmission. In order to reproduce the message at the receiver without any significant distortion, the re-insertion of the carriers for each channel must be done accurately. The accuracy of the carrier supply at the receiver and the transmitter is extremely important. Due to the stringent requirements on the carrier supply, it is economically and practically unfeasible to have a separate supply for each terminal or channel. In practice, the terminal carrier supply is generated from a reference frequency. For the nominal 4 KHz voice channel transmission, a very accurate reference frequency of 4 KHz is generated, and all carrier frequencies for multiplexing are derived by creating its harmonics. This reference frequency is also transmitted to the receiver along with the multiplexed signal. The transmitted reference frequency is called the pilot. Pilots are used as a reference for level control and provide frequency synchronization between the transmitter and the receiver.

The bandpass or the lowpass filters used in practical systems do not have an ideal "box-car" frequency response characteristics: the transition from the pass-band to the stop-band is gradual. Accordingly, some unwanted frequencies are retained during the bandlimitting process. These frequencies may extend into the region of frequencies allocated for other channels, thereby causing overlapping and distortion. In order to reduce this effect, "guard bands" are provided in between the channels. The effect of the guard bands is to increase the transmission medium bandwidth requirement. The term frequency efficiency is sometimes used in connection with frequency

division multiplex systems. Frequency efficiency is the total useful bandwidth divided by the total transmission bandwidth. Frequency efficiency depends on the width of the required guard bands. The width of the guard band, in turn, depends on the slope of the filter response in the transition band. By making the transition band extremely narrow, the required width of the guard band can be reduced. This reduction, however, calls for higher order filters, which require more components, and thus increases the cost of filtering. Thus there is a trade-off between cost and the efficient use of the available bandwidth.

The economics of multiplexing arise from the fact that some of the transmission resources are shared. Instead of having a separate transmission path for each incoming channel, multiplexing allows for the sharing of the same path by many incoming channels. Thus fewer number of transmission lines or cables are needed. Similarly, one repeater is used to amplify signals from different channels. If no multiplexing were used, a separate repeater would be needed for each channel. On a long-haul system, where many repeaters are required, FDM results in a considerable savings.

6-3.2. CCITT MODULATION PLAN. In order to aid the interconnection of large telephone networks on a national and multinational level, the International Consultative Committee for Telephone and Telegraph (CCITT) has recommended a standardized modulation plan. The basic building block of this plan is the nominal 4 KHz voice channel. The CCITT defines the standard group as comprising 12 voice channels occupying the frequency band of 60-108 KHz. Five groups are combined to form a supergroup, equivalent to 60 voice channels. The frequency band occupied by the supergroup extends from 312 to 552 KHz. Thus, five groups can be translated ("re-multiplexed") in frequency to

form the supergroup. Five supergroups are combined to form the basic mastergroup, equivalent to 300 voice channels. The recommended spectrum of the mastergroup extends from 812 to 2044 KHz. Three mastergroups when translated in frequency results in what is called the supermastergroup. The recommended frequencies for the supermastergroup are between 8516 and 12388 KHz.

6-3.3. COMMERCIAL FDM HIERARCHY. The modulation plan used in the U. S. is slightly different from the CCITT recommendations. The "group" and "supergroup" in the Bell System hierarchy are identical to the CCITT recommendations, but the formation of the mastergroup is different. The mastergroup is composed of ten supergroups, equivalent to 600 channels instead of the 300 recommended by the CCITT. Furthermore, the frequency bands occupied by the Bell System mastergroup falls into two different categories: the L600 mastergroup occupies the 60 to 2788 KHz band, while the U600 mastergroup occupies the 564 to 3084 KHz band. Three mastergroups and one supergroup are combined for transmission on the L3 coaxial cable carrier. The L4 system consists of six mastergroups multiplexed to form 3600 channels; L5 and L5A have even larger capacities.

6-3.4. SUB-CARRIER FDM. Just as several voice channels are combined to give a wideband transmission capability, a voice channel can be broken up into sub-channels for low speed data. Each subchannel may be used as though it were a separate line. Telegraph, Telex, and low-speed digital data from teletype terminals, which usually transmit at 75 bit/sec to 1200 bit/sec, can be frequency-division multiplexed and transmitted over a voice grade channel. The frequency division technique in such cases employs tones

within the voice band to transmit data from on point to another. These sets of tones are easily combined in a voice-grade channel. Typically, one pair of tones is used to represent a subchannel to a given terminal. The number of subchannels derived from a voice channel depends on the speed of the terminals being serviced; typically, this number ranges from 2 to 30 on a voice-grade line.

The transmission of up to 600 bits/sec is considered reliable over virtually any telephone connection which is good enough for speech. 1200 bits/sec can also be obtained on a majority of switched connections, and is used extensively on leased circuits. Thus, data from four low speed terminals operating at 300 bits/sec can be frequency-division multiplexed and transmitted over the subchannels derived from a voice channel which can reliably transmit bit rates of up to 1200 bit/sec. Line conditioning becomes necessary for transmission of data rates above 1200 bit/s.

6-4. MULTIPOINT SERVICE; ANALOG HUBS

Multiplexing schemes as discussed above allow many different signals to be combined for transmission over a single channel. Often there is need to send the same signal over several channels. A common example of this situation is the conference call where, in turn, each party to a conference can speak to all the others. To make such a conference distribution one cannot merely connect all the wire pairs together directly. Such a parallel connection would result in impedance mismatches that are intolerable.

In order to provide conferencing with proper termination on each line, special bridges are available commercially. These conference bridges are available with up to six ports. They come in both passive and active configurations. The most efficient passive bridge would be a multiwinding

transformer that matches the impedances with little or no loss. One must remember that if a signal is sent out over two lines only half power, at best, can be sent over each line. Thus a passive bridging network must be followed by an amplifier for each outgoing signal if signal levels are to be maintained. The most common and least expensive hub networks are resistive bridges. These have much higher insertion loss. In fact, the six line hubs have a 20 dB loss. That is, power in on one line leads to one onehundredth of the power out on the other five lines

Since only the outbound power from a conference bridge is to be amplified, the circuits in and out must be four wire if each line is to be able to talk to the other. The required configuration for a three line conference is shown in Figure 6.2.

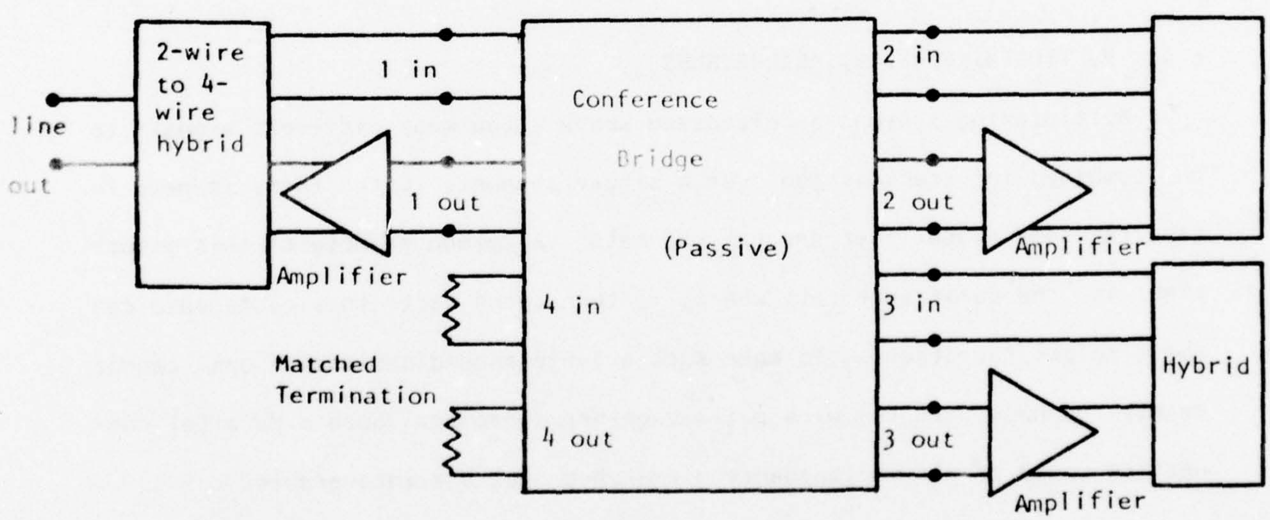


Figure 6.2. Conference circuit with passive bridge.

Commercial active conference bridges are also available. These, in effect, contain the amplifiers along with the passive bridge of Fig. 6.2. They can also be obtained with the hybrids included so that only two wire connections need to be made to the jacks on the box.

Since the bridge must maintain the proper line impedance as well as provide signal distribution, it is important to have all bridge ports properly terminated. In Fig. 6.2 a four-circuit bridge is illustrated for a three circuit conference. The fourth port must be terminated with matched loads for proper bridge operation.

6-5. TIME DIVISION MULTIPLEXING (TDM)

The basic idea in time-division multiplexing is the assignment of a time slot to each incoming channel, during which it can send information without interference from another channel. Time-division multiplexing techniques combine the incoming bit streams into an aggregate bit stream containing the message and the overhead bits (for signaling, synchronizing, etc.).

The allocation of the time slots in the outgoing signal is done in some pre-determined manner by the multiplexer. In theory this allocation can be done in an almost unlimited number of ways. In practice, however, the schemes employed usually fall into one of the four categories: The simplest case is when the incoming channels have the same bit rate. The first possibility is then to combine these streams bit by bit. Such a multiplexer can be considered equivalent to a commutator. No storage or elastic buffers are required in this scheme. A second alternative is to accept characters, represented by a group of pulses, from each incoming channel in turn to form the aggregate bit stream. This is called "character interleaving." This ar-

rangement is clearly more complicated, since the multiplex switch has to halt each time while the entire character is transferred. Some local storage is needed for such a multiplexing scheme. Despite the inherent complexity, this latter scheme is fairly commonly used due to the operational advantage accrued by preserving the code words for each character. The remaining two schemes can be derived from these two cases for incoming signals which are not all at the same bit rate. As an example, consider a 9600 bit/sec incoming channel and four other 2400 bit/sec incoming channels. The multiplexing of these incoming channels can be accomplished using 8 time slots, with alternate time slots allocated for the 9600 bit/sec channel. In a character-by-character multiplexer a similar result can be attained by combining different number of characters from each incoming channel.

It should be evident that the minimum length of the multiplex frame must be a multiple of the lowest common multiple of the incoming bit rates, and hence this sort of arrangement is only practical when there is a simple relationship between these bit rates. Here the case of synchronous multiplexers is assumed; the concepts pertaining to completely asynchronous multiplexer will be dealt with in the next section.

Often characteristics of all incoming lines at the multiplexers are not alike. Furthermore, the effect of ambient conditions on each of these lines may be different. When digital signals are multiplexed on a time-division basis or when a hierarchy in time division multiplexing is desired, this variation in the characteristics can pose problems. The information from incoming channels cannot be directly fed to the selector switch. An intermediate storage is necessary to insure that pulses from the incoming channels are picked up in a proper sequence when a frame is generated. Elastic stores, also called data buffers, are used for intermediate storage. In-

formation pulses arriving at the multiplexer are stored in their respective "cells" in the data buffer. When their turn comes, these pulses in the data buffer are applied to the transmission system, and new incoming pulses are stored in their place. Thus, data buffers constitute an important part of the multiplexer in a time division multiplex system. The variation in propagation time in the medium can cause the slowing or speedup of the arrival of the pulses at the demultiplexer, too. The receiving equipment clocks the incoming pulses at a constant rate. Thus, elastic stores are also needed at the receiver to avoid the overlapping or smearing of pulses.

The synchronization between a multiplexer and a demultiplexer, or between two multiplexers, cannot be established by simply introducing data buffers; the receiver must know <u>when</u> to expect a pulse or a space. This leads to the <u>synchronization problem</u>. It is basically a timing problem, dealing with the establishing of synchronization between the transmitter and the receiver. At the receiver, the information pulses in the frame from the transmitter must be separated and distributed to the outgoing channels. Thus, the identification of pulses for the different destination channels has to be made at the receiver too. This leads to the <u>framing problem</u>.

There are several methods for dealing with the synchronization problem. A master clock can be used to time the entire system. An obvious problem with such a scheme is the inherent vulnerability of the system. A breakdown of the master clock or a transmission link will cause the system to stop functioning. Another method of synchronization is to use very stable clocks at each office using time division multiplexing. Another method which is commonly used in the design of multiplexers is called <u>pulse stuffing</u>. The idea behind pulse stuffing is to have the multiplexer output in bits per second always higher than the sum of the lower speed inputs.

This is accomplished by introducing spurious bits in the output stream that are subsequently withdrawn at the demultiplexer. The information regarding the number and location of the stuffed bits is passed on to the demultiplexer on a separate data line. With the outputs operating at a speed greater than the inputs, no information bits are lost. Pulse stuffing simplifies clock design and also reduces the necessary buffer storage. When this scheme is carried out independently for each multiplexer, the failure of one multiplexer or a line affects only the signals passing through that multiplexer or the line. Thus, pulse stuffing also contributes to system reliability. Another simple and commonly used method for synchronization is to use a phase locked loop (PLL) in the transmission medium modem to recover the signal (carrier) used for synchronization purpose.

The framing problem is dealt with by either affixing a unique code with each code word or by utilizing the intrinsic characteristics of the signal itself. There are two methods under the former approach: First, the added digit framing, in which a position is reserved for the signaling information; second, the robbed digit framing, in which the least significant bits of the actual information bearing code words are used to transmit the signaling information on a regular but infrequent basis. Strategies based on using the characteristics of the signal itself depend upon the statistical behavior of the digit positions in the pulse stream. This method is called statistical framing. Any one of these methods can be used to identify the beginning of the frame. For the identification of the individual components of the frame, the added digit framing strategy is commonly used. For example, the D1 system for time division multiplexing of voice channels uses 7 bits to represent a code word and an additional bit for signaling information. Twenty four channels are multiplexed in a frame, and a framing bit is

used to identify the beginning of the frame. Thus, each frame contains 193 bits. The D-2 and later systems keep the 193 bit frame while adding one bit (now 8) to the code word. In every sixth frame the least significant bit of the code word is replaced by a signaling bit.

6-5.1. ASYNCHRONOUS TIME-DIVISION MULTIPLEXING. Where synchronous time-division multiplexing between multiple original data sources is difficult to achieve, asynchronous time-division multiplexing may be used. Asynchronous multiplexer systems store the data input signal from each incoming channel temporarily until a prescribed number of data bits (block of data) becomes available for a particular input source. An independent clock signal for each incoming channel from a standard bit synchronizer is used to accomplish this initial storage. Before transmitting the block of input data, a header ("label"), made up of a unique bit pattern for each input source, is appended to the information bit stream. In this way, the data sources are multiplexed as blocks of information rather than bit-by-bit or character-by-character multiplexing, as is the case of synchronous multiplexers. The unique bit pattern for identification of each channel permits proper demultiplexing at the receiver.

It should be noted that such a multiplexing system can also be operated in a condition where useful overhead data, such as coding check bits, low priority data, etc., can be transmitted when no incoming channel has sufficient data to be processed. Also, since data from incoming channels is shifted into the multiplexer by their own clock signals, the system operation is not affected by data rate variations.

The statistical characteristics of the generation of data by remote terminal users has led to new forms of asynchronous time-division multiplex-

ing. A case in point is the ALOHA system. Consider a number of widely separated users each wanting to access a computer by transmitting their data over a high speed transmission line. In a commonly encountered situation, the average time between data packet generated from a single user is much greater than the time needed to transmit the single data packet. The following asynchronous scheme is then used: Each user station (remote terminal) has a buffer storage which is used to store the incoming data. When a certain length of data has been stored a header containing address, control and parity information is appended to this data to form a "packet." This packet is then transmitted to the central station. Thus, each user at a remote terminal transmits packets of data to the central office in a random and completely asynchronized manner. Such a channel is also referred to as a random access channel. The central office then acknowledges if the received packet is error-free or not. If no acknowledgement is received at the transmitting station, the packet is automatically retransmitted. Two types of errors can occur in such a scheme: random noise errors and errors caused by interference with a packet transmitted by another remote terminal. The first type of errors are not unusual in transmission. It is the second type of errors that are unique to this type of an asynchronous multiplexing. These errors grow as the number of users becomes large. These interference errors usually limit the number of users and the amount of data that can be transmitted over the random access channel to less than 15% of the full synchronous capacity.

6.5.2. TDM HIERARCHY. Just as in frequency division multiplexing, a hierarchy in time-division multiplexing system is also present. On a typical voice-grade channel, data of up to 1200 bits/s can be transmitted

without line conditioning. Transmission of data above 1200 bits/sec, in particular, 2400 and 4800 bits/s, requires line conditioning. Based on the low-speed teletype and other preferred data rates, a hierarchy of the form $Cx75x2^N$, N = 1,2,...6, bit rates has become standard. Thus, four console terminals operating at 300 bits/s (C = 1, N = 2) can be multiplexed for transmission over a 1200 bit/sec circuit.

6-6. STATISTICAL TIME-DIVISION MULTIPLEXING

The development and availability of low-cost microprocessors has led to a new form of time-division multiplexing, called statistical time-division multiplexing (STDM). This technique dynamically allocates the available bandwidth of a high speed transmission line amongst the lower-speed terminals connected to it. In both FDM and conventional TDM, a fixed percentage of the line bandwidth is permanently allocated to each user, usually on the basis of transmission speed, and regardless of whether the user terminal is active. Thus, this fixed portion of the bandwidth (bit rate) is wasted when the terminal is inactive. In STDM, this is not the case; the user will only be provided with a channel when there is data to send. During idle periods, this same bandwidth is allocated to another user. Thus, the basic assumption behind the use and effectiveness of STDM technique is that not all terminals connected to a line will require service at a given time; that is, the aggregate bit rate requirement is less than the sum of the terminals. To use a simplistic example, consider a 4.8 Kbit/sec circuit. A time-division multiplexer will use this circuit among four 1200 bit/sec users. If each user is active less than 50 percent of the time, a statistical time-division multiplexer could support eight terminals on that same 4.8 Kbit/sec circuit.

Since the capacity allocation in FDM/TDM techniques is fixed, the subscriber terminals cannot interfere with each other and, consequently, the delay (an important system performance measure) incurred at a terminal is not influenced by the number of active terminals. STDM, on the other hand, can be sensitive to the number of users desiring to transmit data. If many channels are simultaneously active, and the sum of this activity approaches or exceeds the high-speed data rate circuit, the transmission delays can grow as more user data is queued in the STDM transmitter awaiting transmission on the circuit. The number of channels supported by the multiplexer will depend on the maximum amount of delay that can be tolerated.

The potential delay problem notwithstanding, STDM techniques are more efficient than FDM/TDM techniques in relation to how many users can share a circuit. Other statistical mux benefits include an error-protected protocol on the high-speed links. Also, data errors arising due to transients in telephone lines will not be seen, because the device will retransmit the block in error.

Statistical multiplexers use a frame structure to transmit data from various incoming channels. Since the channels are allocated dynamically, the data packets from users must be addressed in some pre-determined fashion for proper demultiplexing at the receiver. The length of the frames used depends on the amount of user activity and the overhead bits appended for control. The frame length used is an important performance parameter.

6-7. LINE CONCENTRATION.

The cost of laying cables in a telephone network can get extremely high if proper measures to reduce the amount of cable used are not taken. In order to reduce this cost, and concomitantly the cost of long-distance line

rental, a number of techniques have been devised for lowering the number of required cables for serving the needs of a particular area. Consider, for example, the use of remote terminals in city A connected to a central computer in city B. Clearly, it would be disadvantageous for the subscribers to pay for the long-distance telephone calls everytime they need to use the computer. To overcome this, a concentrator (typically a small processor) can be installed in city A. The remote terminals would be connected to the concentrator, and the concentrator, in turn, is connected to the computer by a high speed line. The subscribers now have to pay for their local calls, while the organization providing the computer services pay for the high speed line rental (recovering the cost indirectly from the subscribers). This arrangement is shown in Figure 6-3.

The concentrator acts as a message collector for the subscribers in the area. Messages are stored in the concentrator's storage for each user until a signal is sent to indicate the end of the message. The complete message, along with the terminal identification information, is sent to the computer. Thus the concentrator acts as a message-switching center and operates on a store-and-forward principle. The output from the computer are sent to the appropriate concentrator, and distributed to correct remote terminal by the concentrator.

The description of concentration techniques above has been for computer networks. Line concentration techniques are also used in telephone networks. For example, if two cities each have 100,000 telephone subscribers who might call each other, there are not 100,000 channels between the cities. The number of actual channels may be a few hundred between these cities. Such an arrangement works because all the 100,000 subscribers do not want to talk to each other at once. Although not done universally, a con-

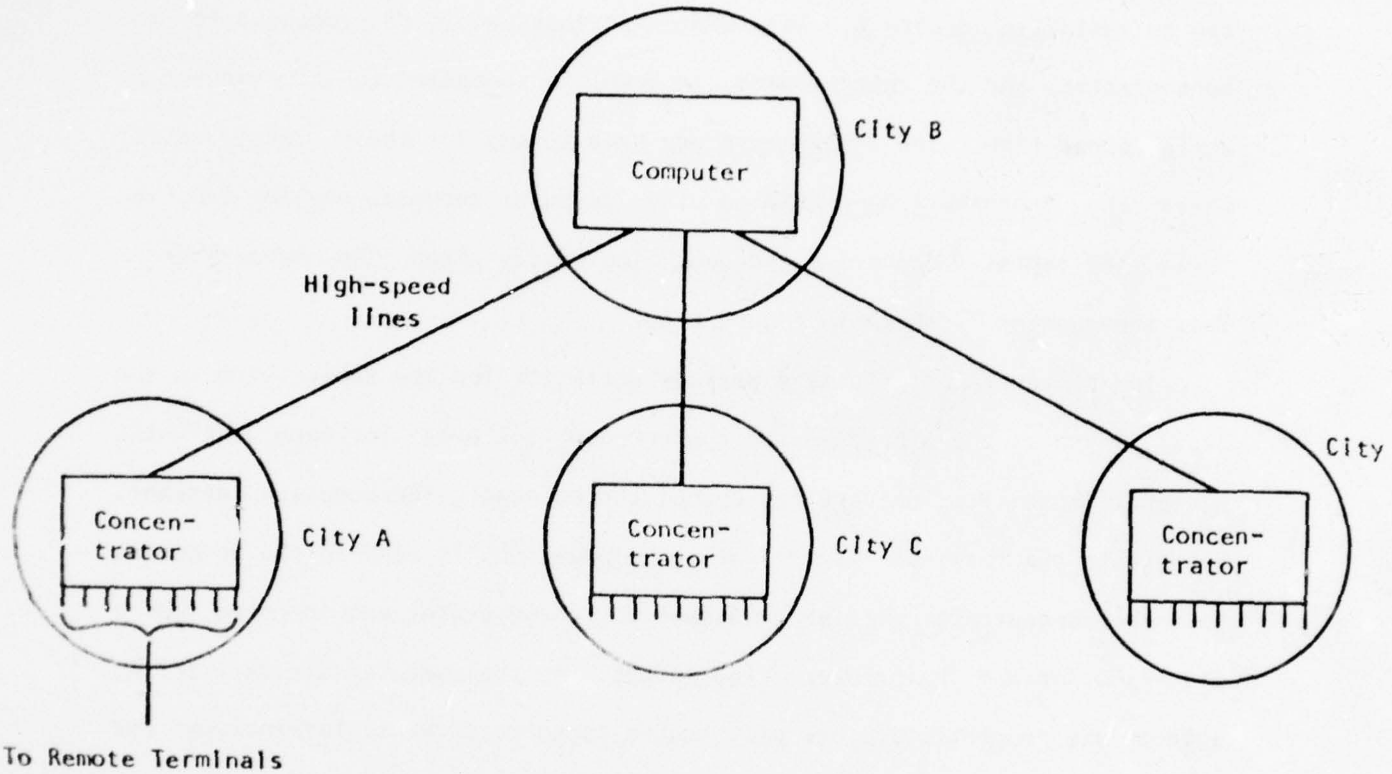


Figure 6-3. An Example of Line Concentration

centrator is often used to reduce the number of lines on the subscriber side of the dial central office (DCO).

Digital speech interpolation techniques, abbreviated DSI, can be used to reduce the number of transmission channels required by making use of the speech statistics of telephone callers. In a typical telephone conversation, one party talks while the other party listens. Thus, on an average, any one-way path is in use only 40% (due to pauses, etc.) of the total time. In other words, for 100 talkers only about 40 on average will be talking simultaneously. Digital speech interpolation techniques aim to use less than 100 transmission channels to transmit 100 incoming channels. Time Assignment Speech Interpolation (TASI) accomplishes this by assigning transmission channels on the basis of speech activity on the channel. The TASI equipment is designed to detect a user's speech activity and assign him a channel within a very short time (in the millisecond range). The channel is retained by the user as long as the speech activity continues; as soon as the activity stops, and if the channel is required for another user, it will be taken away. The cost for TASI equipment is offset by the savings in the amount of cable required. Clearly, such an arrangement is economical on long transmission channels. On shorter channels, the cost of TASI equipment may be higher than the savings derived from using less cable.

6-8. MULTIPOINT NETWORK.

Another technique used for sharing a high-speed line is where a number of devices (terminals and/or concentrators) are connected to the same line. Such a system arrangement is known as a multidrop network, multidrop line, or multipoint network. Figure 6-4 illustrates a multidrop network arrangement. With a multidrop network, all the messages to and from the terminals

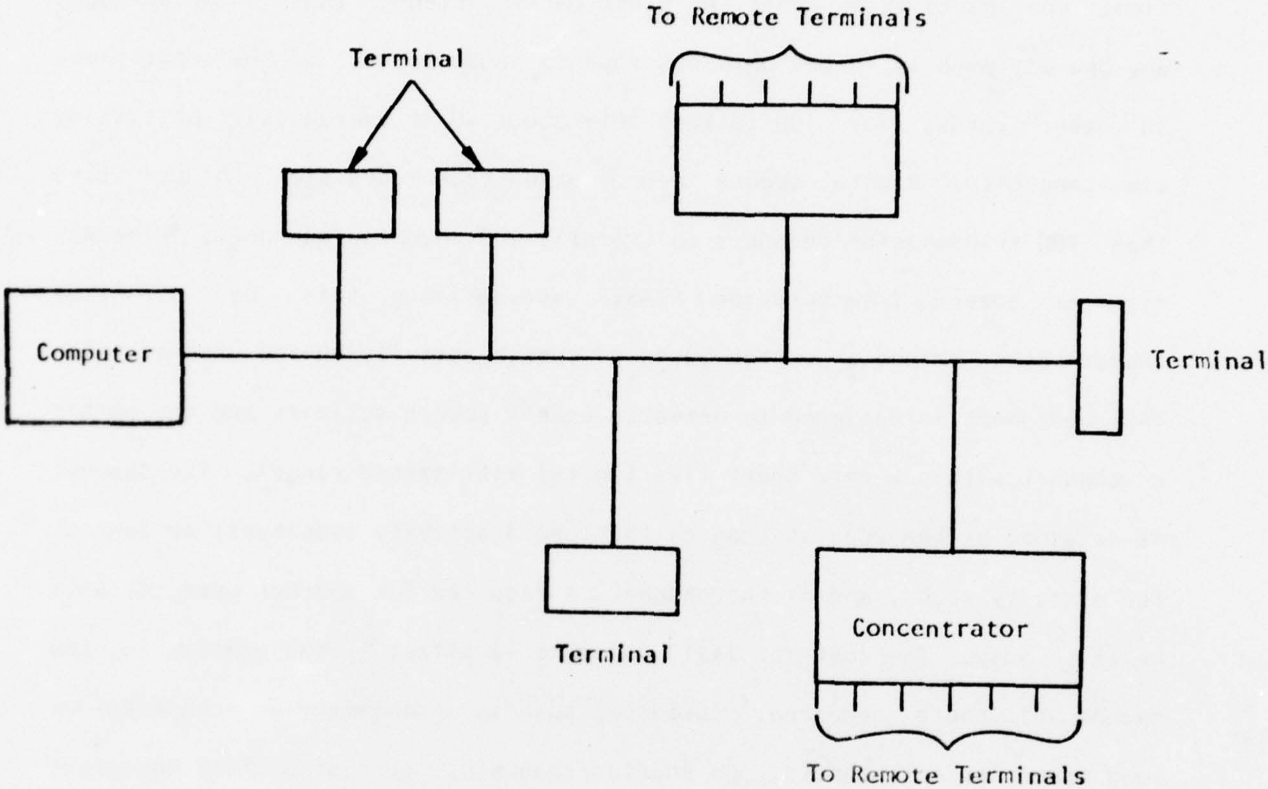


Figure 6-4. Illustration of a Multidrop Network.

are under the control of the central computer (or its communication subsystem). These messages are processed one at a time. Messages from the central computer are affixed with a select message (header), which contains the address of the device on the multidrop line to which the message is directed. All devices except the selected one ignore the select message. On some multidrop systems it is possible to send data to more than one device. If synchronous transmission with handshaking is used, the selected device upon receipt of the select message may acknowledge it; otherwise, it simply accepts the message following the select message.

The receiving of the messages from the devices on a multidrop line is also controlled by the central computer. The computer sends out a polling message with a particular device address. If the selected device has no message to transmit to the computer, a new device is selected for polling by the computer. The devices transmit their messages only when they receive the polling message. The sequence of polling is based on a polling table in the central computer. By placing a given device in the polling table at a number of points, an arrangement to handle devices of varying speeds with varying relative priorities can be established. The polling of devices is the basis of all multiplexing; the multidrop network concept is simply software multiplexing.

6-8.1. DATAHUB. Datahub is the device used for interconnecting private-line telegraph and data circuits and equipment in multipoint configurations. These sets can be installed on customer premises or in a central offices. It is the digital equivalent of the Analog Hub (conference Bridge) discussed in 6-4. As with the multidrop network, headers and/or polling techniques are used.

6-9. PULSE CODE MODULATION (PCM)/TDM

Time division multiplexing can also be used for transmitting voice on one transmission path by dividing the time domain into slots. The messages from different talkers are interspersed in time as they are propagated over the transmission medium. N input bandlimited signals are sequentially sampled (commutated) at the transmitter. The rate of sampling is determined by the sampling theorem: if the incoming signals are bandlimited in W Hertz, then the required number of samples is at least 2W samples per second; that is, the commutator makes one complete cycle in less than 1/2W seconds, obtaining one sample from each incoming signal. The output from the sampler is therefore a pulse amplitude wave containing the individual message samples periodically interspersed in time. Samples from the same source are separated by less than 1/2W seconds, while the distance between two pulses in the PAM wave is the sampling time, $T_s (\leq 1/2W)$, divided by the number of incoming channels. Therefore, for N incoming channels, the spacing between any two pulses is T_s/N seconds. A set of pulses consisting of one sample pulse from each incoming channel is called a <u>frame</u> and the duration of the frame is called the <u>frame length</u>. Thus, in accordance with the sampling theorem, the frame length must be less than or equal to $1/2f_T$ seconds, where f_T is the highest frequency content of the incoming channels.

The pulse amplitude wave from the commutator is not suitable for transmission outside of the telephone plant. Some additional processing of the pulses is needed so that transmission over a long- or short- haul system can be accomplished without introducing any appreciable distortion. Each sample is quantized and coded before transmission. This procedure is called pulse code modulation (PCM) transmission.

At the receiver, the decoder generates a pulse amplitude wave in accordance with the information contained in each received code word. A rotary switch, called the distributor or decommutator, separates the sample pulses from the PAM wave and distributes them to their destination channels. An interpolation filter is used to reconstruct the message at each channel.

The T1-carrier system is a first-generation pulse code modulation (PCM)/TDM system. In common usage, distinction between the line and the multiplex terminal is made: the line is referred to as the <u>T1 repeatered line</u> and the multiplex terminal is referred to as the <u>D1</u> (<u>or</u> <u>D2</u>, <u>D3</u>, <u>etc</u>.) <u>Channel Bank</u>. It is in the channel bank that the basic signal processing--like sampling, quantizing, etc.--involved in forming the PCM signal takes place. It is also the bank that controls system timing.

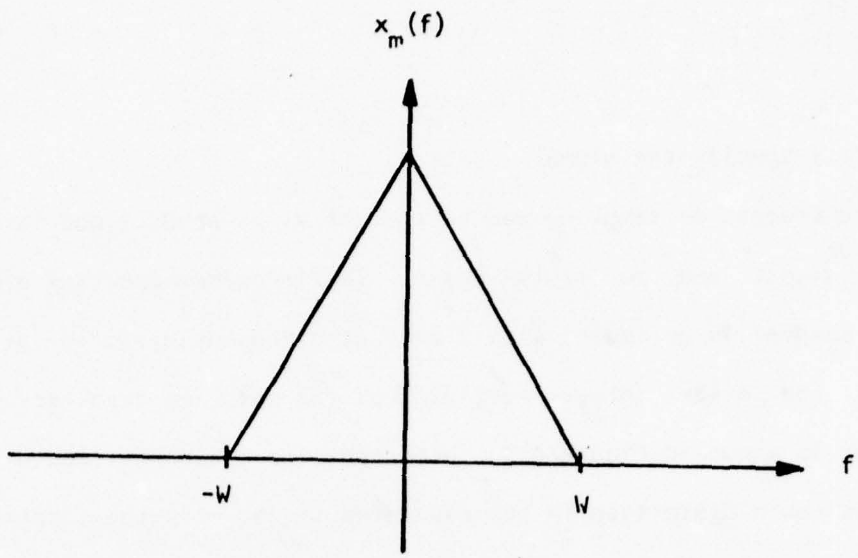
The analog signal is first converted into discrete pulses in time (PAM wave), followed by the discretization (quantization) of the pulse amplitudes. These pulses are then coded using logic circuitry. The codes are expressed as a series of identical pulses, or spaces. A pulse indicates a binary "1" and a space indicates a binary "0". The series of pulses and spaces that defines one quantized sample from one channel makes up a PCM word. The D channel bank combines 24 voice channels to make up a "frame," containing one eight-bit word from each channel. Thus, 192 bits makes up a "frame" in the T1-carrier system. For synchronizing purposes, the beginning and end of a frame must be identified. To accomplish this, a 193rd time slot is inserted in each frame to provide timing information. Thus, the T1-repeatered line rate is 193 bits/frame x 8000 frames/sec = 1.544 Mb/s.

So far an overview of the signal processing involved in PCM T1-carrier system has been presented. Some of these steps are now discussed in more detail.

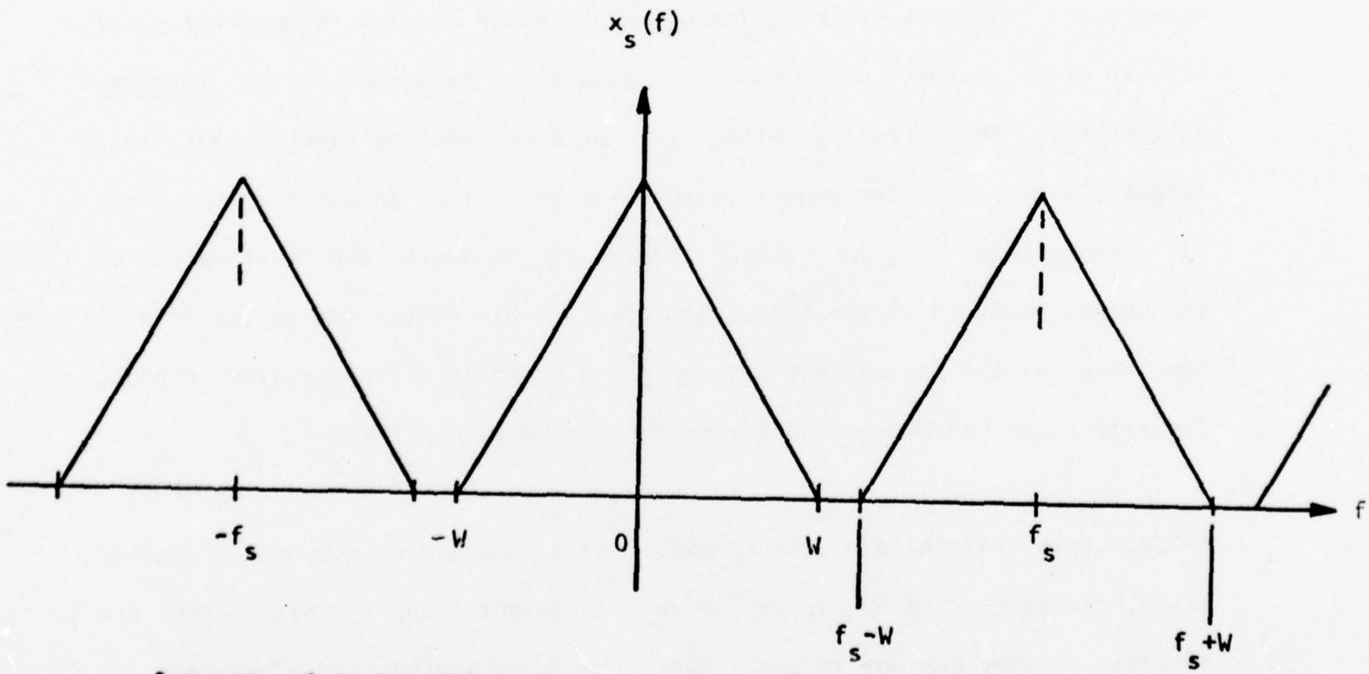
6-9.1. SAMPLING. All telephone conversations and most of the other signals encountered in transmission systems have a finite time duration, that is, they are present for a finite amount of time. Such signals are called time-limited signals. All physical signals fall into this category. According to theorems in communication theory, a time-limited signal cannot be bandlimited. The frequency spectrum of all time-limited signal extends over all frequencies. However, virtually all of the signal energy of the common signals of human communication is concentrated in a finite frequency band. Because of this fact, it is appropriate to talk of signals which are both bandlimited and time-limited at the same time.

Due to the bandlimited nature of signals, the change in signal amplitude in a small period of time is also limited (finite). By specifying the signal amplitude at discrete time steps, it is possible to specify the signal waveform. The process of generating a discrete signal (sequence of numbers) from a continuous time signal is called sampling. The pulses at each discrete time is called a sample and its amplitude denotes the amplitude of the analog signal at that time. When the sampling interval is uniform, the process leads to what is called uniform sampling.

The sampling interval, T, cannot be selected arbitrarily. If the sampling rate is too slow, some of the information will be lost and it will be impossible to reconstruct the original message. If the sampling rate is too high, the amount of communications assets required to transmit the information will become unnecessarily high. The minimum sampling rate, called the Nyquist rate, is determined on the basis of the sampling theorem. According to the sampling theorem, the sampling rate must be at least twice the highest significant frequency of the bandlimited message. Thus, for a signal bandlimited in f_T Hertz, at least $2f_T$ values per second are needed to



Bandlimited Message Spectrum



Spectrum of an ideally sampled message.

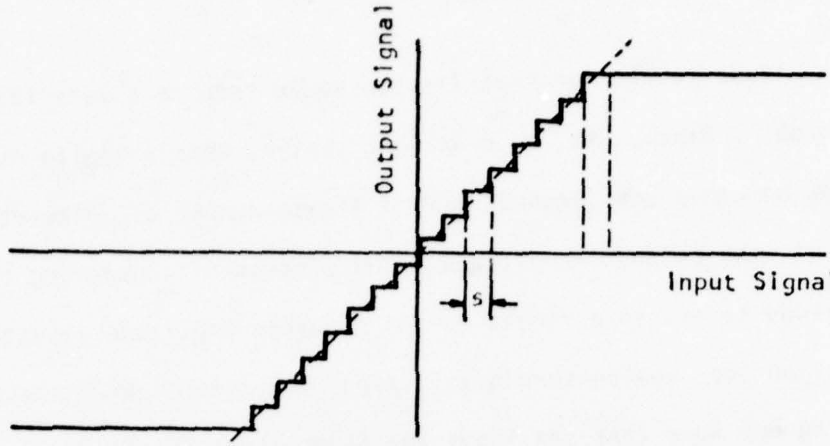
Figure 6-5.

completely specify the signal.

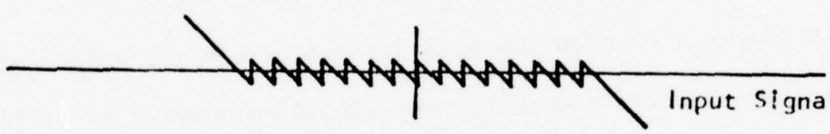
The process of sampling can be thought of as product modulation of an analog signal and an impulse train. The frequency spectrum of the output from a sampler is periodic, with a pair of sidebands centered at d-c, f_S, $2f_S$,... and other integer multiples of the sampling frequency f_S. Such a spectrum is shown in Figure 6-5. Note that the sidebands would overlap - and thus cause distortion in the recovered signal - if the sampling frequency were not twice the value of the highest frequency in the message.

In order to bandlimit the input signal before sampling, a low-pass filter is used. Since all low-pass filters have a finite slope in the stop-band, some of the higher frequencies are not completely removed from the input signal. These higher frequencies are manifested in the form of overlapping of sidebands in the spectrum of the output from the sampler. This overlapping effect is called aliasing which, after message reconstruction at the receiver, constitutes what is generally referred to as foldover distortion. The aliasing effect can be alleviated by sampling at a rate slightly higher than the Nyquist rate. Another effect encountered in digital transmission as a consequence of sampling is due to the finite width of the pulses involved in the sampling process. This effect due to the finite "aperture" of the pulses causes frequency distortion of the original signal. The effect can be compensated for by the use of linear filters.

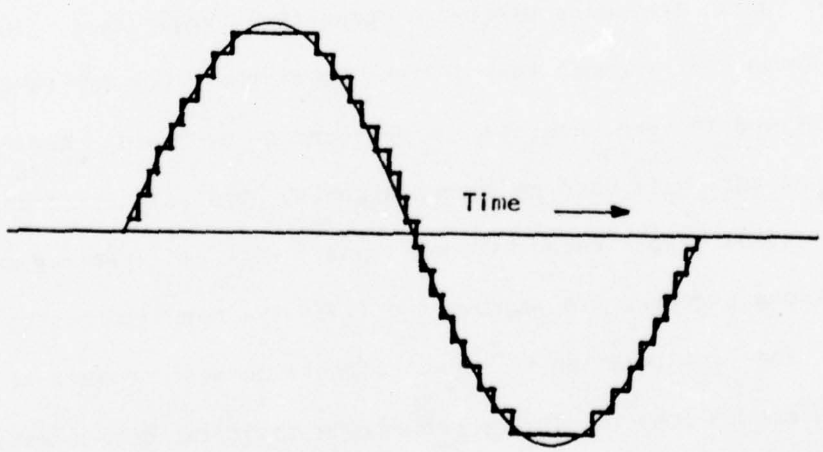
6-9.2. **QUANTIZATION (A/D CONVERSION).** The output of the sampler is a pulse train, called the PAM (Pulse-amplitude modulation) signal. These pulses are discrete in time but not in amplitude. The input analog signal has an infinite number of amplitude levels. The sampled version of the waveform also has a continuous set of possible amplitude levels. To encode each one of



(A) Uniform codec transfer characteristic



(B) Error characteristic



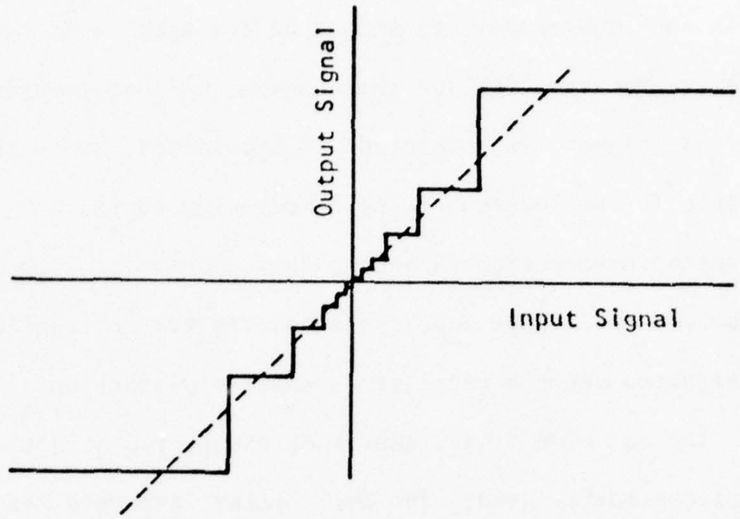
(C) Quantized full load sine wave

Figure 6-6. Characteristics of a uniform codec.

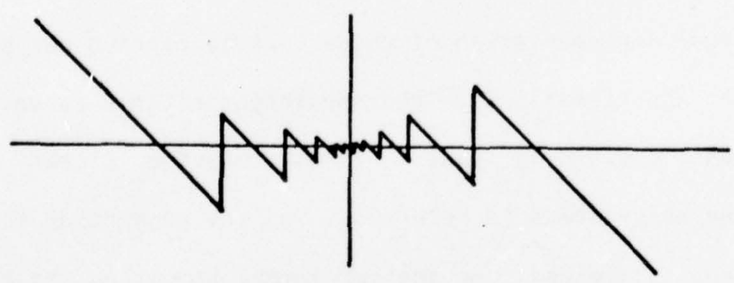
these amplitude levels for transmission would require a very large number of code words. Since, in a practical system, only a finite number of code words can be coded and decoded, only a finite number of discrete amplitude levels of the pulses are allowed. The process of converting the continuum of amplitude levels to a finite set of discrete amplitude levels is called quantization or analog-to-digital (A/D) conversion. When the quantization is carried out such that the input amplitude range is divided into steps of equal width so that the output levels are also equally spaced, the process is referred to as linear or uniform quantization. The input-output characteristics of such a quantizer are shown in Figure 6-6.

The difference between the quantized value and the actual value of the sample leads to an error when the message is reconstructed at the receiver. This type of error is the major source of imperfection in PCM transmission, and is referred to as <u>quantization distortion</u>.

The distribution of amplitude levels in speech signals is not uniform: small amplitudes are more likely to occur than larger ones. Furthermore, since the human ear responds more or less logarithmically for large signals, small variations in large amplitude signal cannot be heard. Accordingly, if uniform quantization is used on speech signals, the signal-to-quantization distortion (S/D) experienced by weak signals will be excessive compared to that for strong signals. To improve the fidelity, nonuniform step sizes are chosen in the quantization of speech signals on most commercial carriers. Two methods can be used to achieve nonuniform quantization: first, a nonuniform quantizer, whose transfer characteristics are shown in Figure 6-7, can be used; second, the samples of the PAM wave can be predistorted before uniform quantization. The predistortion process used in speech transmission involves the compression of the samples at the transmitter. At the re-



(a) Nonuniform codec transfer characteristic



(b) Error characteristic

Figure 6-7. Characteristics of a nonuniform codec.

ceiver, the inverse process of expanding is used. This process of compression and subsequent expanding of the samples is called companding. The transfer characteristics for compression, uniform quantization, and expanding of a signal is depicted in Figure 6-8. Note that the expander characteristic is the inverse of the compression curve.

The type of compression-expanding characteristics used depends on the distribution of the signal amplitudes and the type of application involved. In speech signals, where a constant signal-to-distortion ratio is to be maintained for a wide signal amplitude range, two slightly different companding laws are mostly used. The Bell system standard is called μ-law. The CCITT is called A-law. Both have compression characteristics that approximates a logarithmic curve for large amplitude signals. The resulting signal-to-distortion characteristics, which depend on the value of μ or A chosen in the respective law,is nearly constant for a wide signal dynamic range. Hyperbolic or other nonlinear transfer characteristics can be used for compression, too.

The actual implementation of these laws is carried out by using piecewise linear approximation to the smooth logarithmic curve. Such laws are called segment companding laws. Each piecewise linear region in the transfer characteristics is referred to as the segment or chord. Each segment, in turn, is divided into smaller intervals called the quantizing levels. The 13 segment A-law and the 15-segment μ-law are the most commonly used companding laws. In actuality these laws use 16 segments - eight segments for positive amplitudes and eight segments for negative amplitudes - to approximate the logarithmic curve. In the 13-segment A-law the four middle segments (two segments on either side of the origin) are collinear, thus resulting in 13-segments. In the 15-segment μ-law, the two middle segments

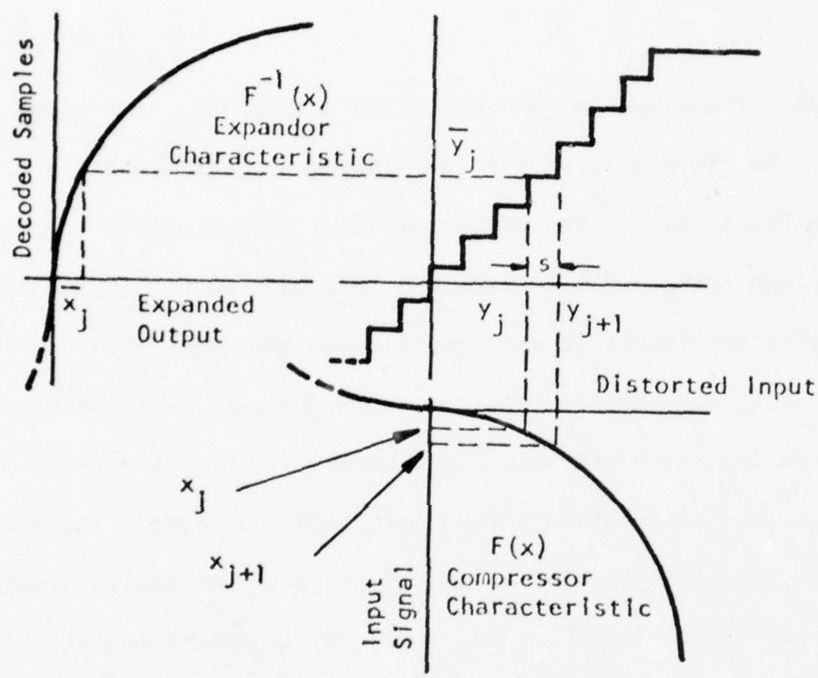


Figure 6-8. Uniform Quantization with Companding

are collinear. Each segment is further divided into 16 quantizing levels. Segment laws, in comparison with "smooth" laws, have the advantage that certain digital processing operation such as filtering, conferencing, echo suppression, net loss adjustments, equalization, companding law conversion, etc. can be performed efficiently with digital circuits.

6-9.3. CODING. The value of each quantized sample is transmitted to the receiver in the form of a code word. Thus coding is closely connected to quantizing in PCM systems. The number of quantization levels chosen determine the minimum number of bits required in a code word. If the difference between quantization levels is made small, then the quantization distortion becomes small too, but the required bit rate increases. If the quantization is made too coarse, the distortion might become intolerable. This underscores the trade-off involved between bit rate requirement and distortion, similar to the trade-off between noise and bandwidth in analog systems.

When either the 13-segment A-law or the 15-segment μ-law is used during the quantizing of the PAM wave, the value of each sample can be coded in terms of the segment to which it belongs and the quantizing level within the segment. Since both laws have basically 16 segments and 16 levels within each segment, thus giving 256 possible amplitude levels, a total of 8 bits ($2^8 = 256$) are required to code each sample. In the μ255-law, each code word is represented as (sabcwxyz) bits, where the (wxyz) bits denote the quantization level within the segment number represented by the (abc)-bits, and the s-bit denotes the polarity of the sample. The (sabc)-bits which basically represent the segment information are called the "characteristic" bits of the word, and the (wxyz)-bits are called the "mantissa" bits. Figure 6-9 shows the coding and quantizing using A-law companding.

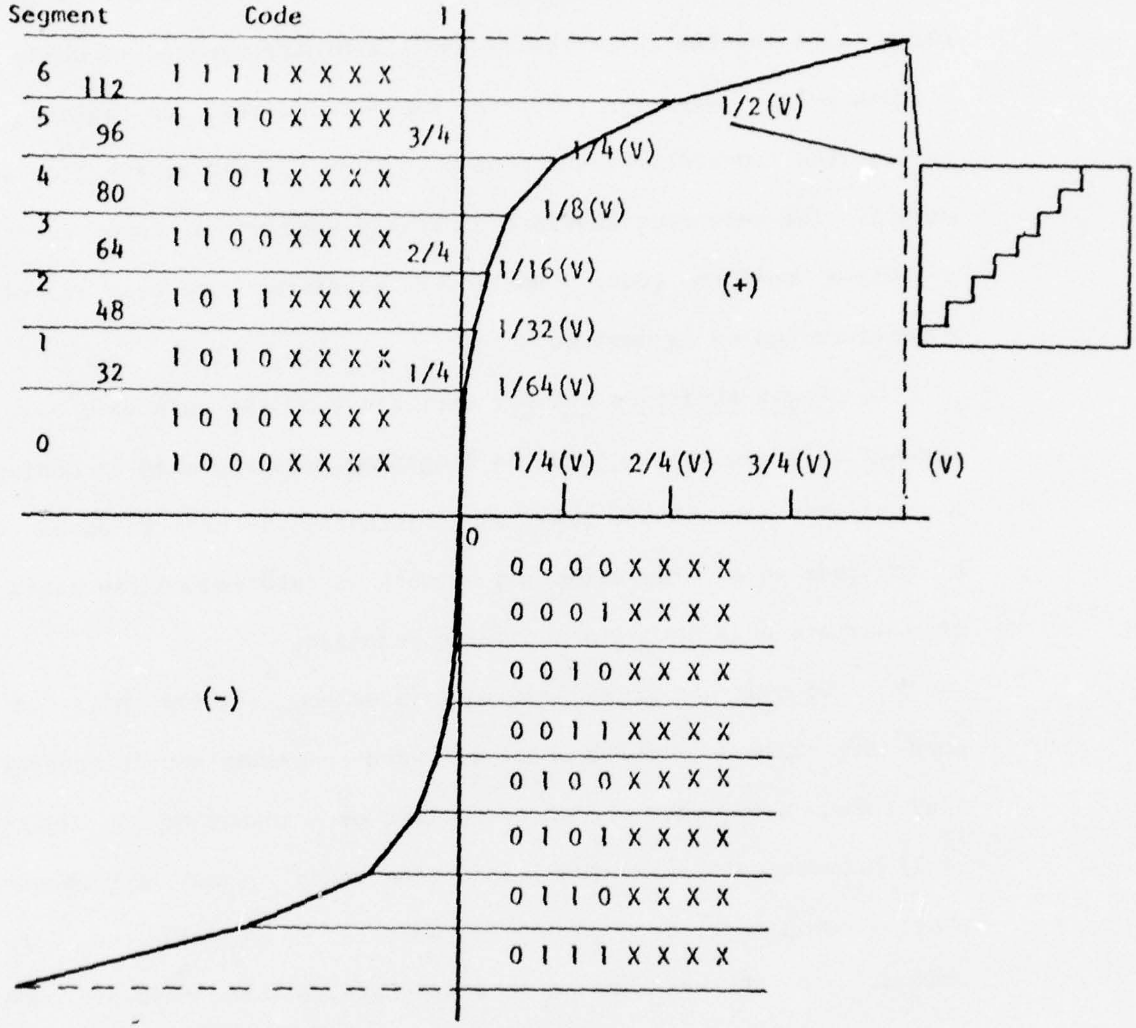


Figure 6-9. 13 Segment A-law: for companding and coding.

The way in which a code word is formed determines the coding method. There are three such methods:

1. Level-at-a-time Coding: In this method a reference waveform is compared with the PAM sample value while a binary counter is being advanced by a clock pulse. When the reference waveform exceeds or equals the sample value, the contents of the binary counter is used as the code word for the sample. The reference waveform generally used is a ramp function, which yields a uniform code. Nonlinear waveforms are used as reference when nonuniform coding is desired.

2. Digit-at-a-time Coding: Each digit of the code word is determined sequentially in this method. A weighting network, used in conjunction with a comparator circuit and some logic circuitry, is used to obtain the digits of the code word. The weighting network is selected on the basis of whether a uniform or a nonuniform code is required.

3. Word-at-a-time Coding: In this method, all the bits of the code word are formed simultaneously. Word-at-a-time coders generally use the Gray code. Gray codes are characterized by a change of one digit (distance of 1) between each pair of adjacent code words. A multiple threshold coder, a beam coding tube, or some other scheme can be used for the word-at-a-time coding. In the beam coding tube approach, all the possible code words are stored on a code plate in the form of holes or no holes. Depending on the PAM sample value, a beam is deflected on the plate and the associated code word separated. In the multiple threshold coder, logical threshold circuits are used to sense the PAM sample voltage to produce the code word. Word-at-a-time coding is the fastest amongst the three methods.

6-9.4. DIFFERENTIAL PULSE CODE MODULATION (DPCM). Speech and video signals

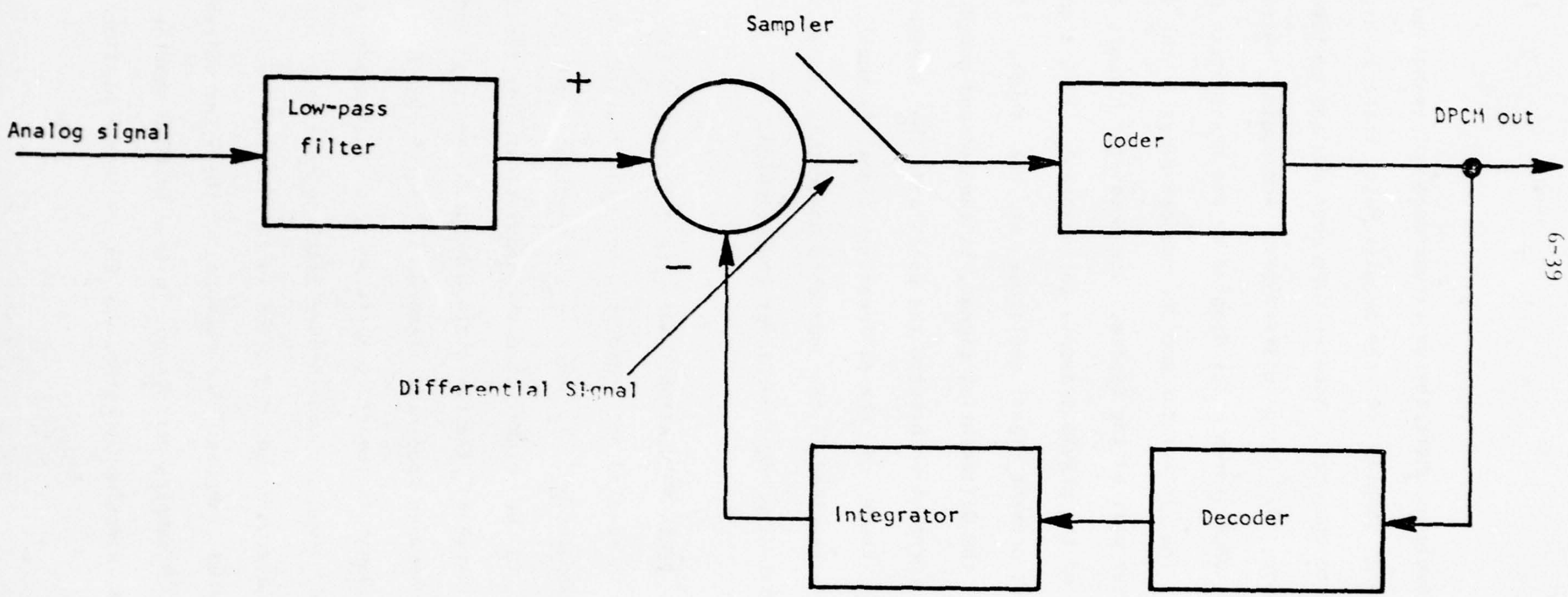


Figure 6-10. DPCM block diagram.

have waveforms where the amplitude does not change quickly. When such signals are sampled at the Nyquist rate, there is high correlation from one sample to the next. Schemes have been devised to lower the bit rate requirements by using differential PCM coding, instead of straight PCM. A typical DPCM system block diagram at the transmitter end is shown in Figure 6-10. The integrator and the decoder block in the feedback path forms the predictor part of the system. The predictor outputs a signal based on the value of the previous sample (or samples). This signal value is subtracted from the present signal amplitude at the input. The resulting signal, called the differential signal, is sampled and coded for transmission. Due to the correlation between the past and the present sample values, the dynamic range of the differential signal is small. Thus, fewer number of bits are required in the code word under this scheme. This results in a lower bit rate requirement for transmission.

6-9.5. DELTA MODULATION. When the differential signal, as marked in Figure 6-10, is sampled and coded into one binary digit, the DPCM scheme is called delta modulation. In general, the sampling rate in delta modulation is chosen to be higher than the Nyquist rate for the signal being processed. This "increases" the correlation between samples, and thus the amplitude change between sequential samples is small. One bit representation for this small change is therefore sufficient for subsequent reconstruction of the message. Even with the higher sampling rate used in delta modulation, there can be a significant reduction in the bit rate requirements for moderately correlated samples. For example, if the 4 KHz voiceband message is sampled at 32,000 samples per second, instead of 8000 samples per second as dictated by the sampling theorem, and the differential signal coded into 1 bit/per

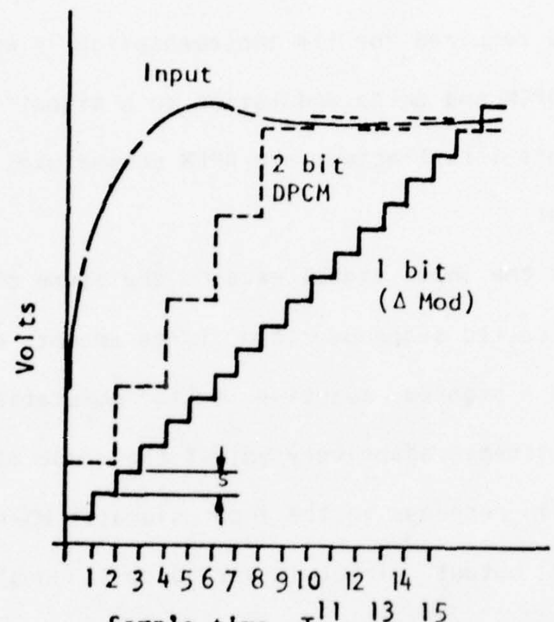


Figure 6-11. DPCM and Delta modulation response to an input signal.

sample, the resulting bit rate is 32 K bits/second. For straight PCM using 8-bit codewords, the required bit rate is 64 K bits/second. The price paid for the lowered bit rate requirement is the degradation in the quality (fidelity) of the received message. Whenever high quality transmission is required, straight PCM has advantage over delta modulation in terms of bandwidth requirements. On the other hand, delta modulation has the advantage that the hardware required for its implementation is much simpler.

The response of DPCM and delta modulation to a signal is illustrated in Figure 6-11. In this illustration, the DPCM scheme uses two bits to code the differential signal.

When the slope of the input signal exceeds the slope of the delta modulation output signal, called slope overload, large amounts of distortion can occur. To combat such a problem, adaptive delta modulation schemes have been devised. These schemes adaptively adjust the slope of the delta modulation output signal in response to the input signal. When the slope of the delta modulation output signal is varied continuously, the resulting scheme is called continuous variable slope delta modulation, abbreviated CVSD. The implementaticn of CVSD can be carried out in two ways: analog or digital. Performance of CVSD depends to some extent on the type of implementation. These schemes are often called "discrete adaptive DM" and "continuous adaptive DM." A block diagram for the implementation of digital CVSD is shown in Figure 6-12. In straight delta modulation the quantum step sizes were fixed. In the scheme shown in Figure 6-12, the switch control chooses a gain G_i by which to increase the quantum steps. This choice is dictated by a logical decision process base on observations of the sequence of pulses leaving the quantizer. For example, when slope overload occurs, thus causing distortion, the quantizer output is a series of pulses of the

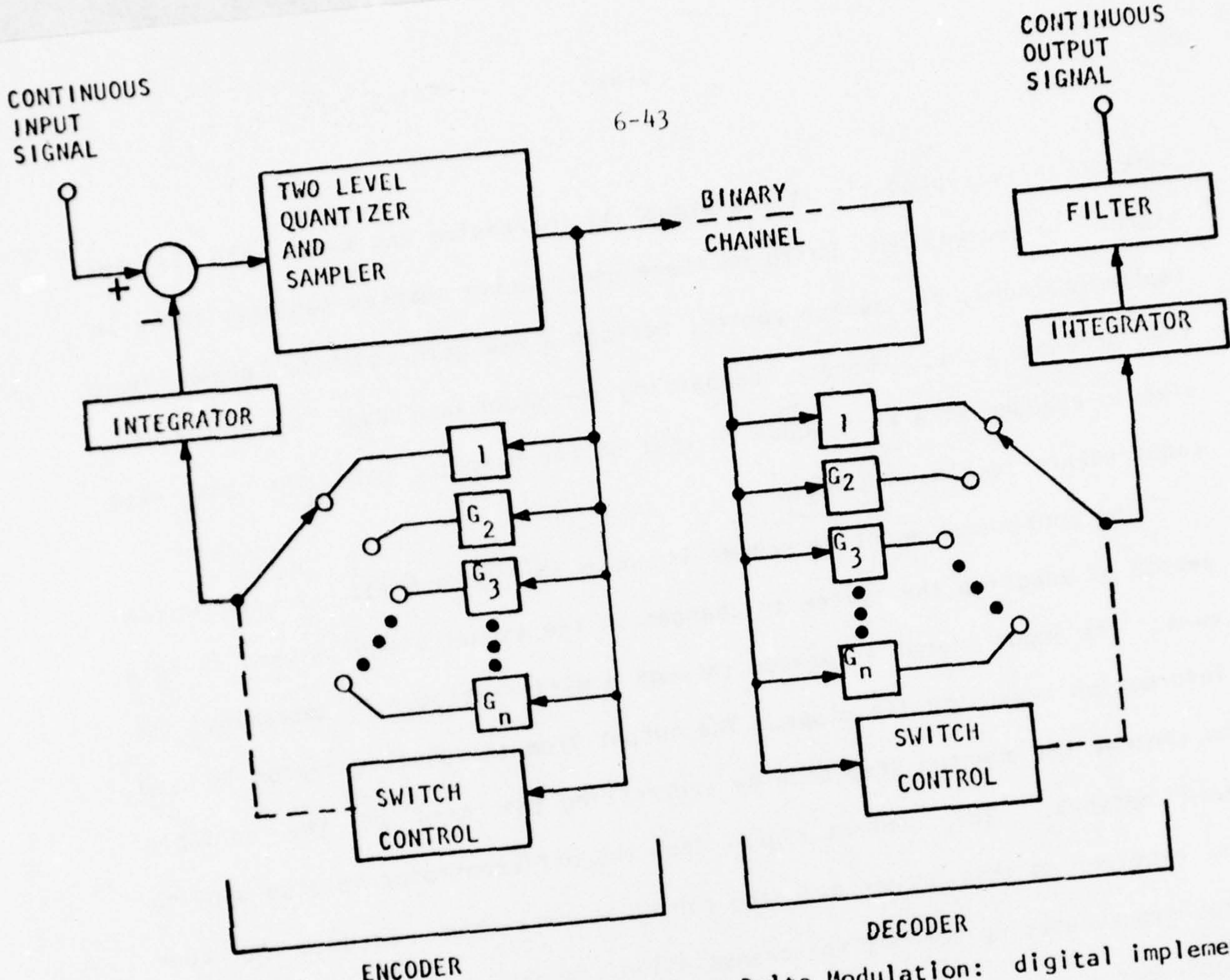


Figure 6-12. Continuous Variable Slope Delta Modulation: digital implementation

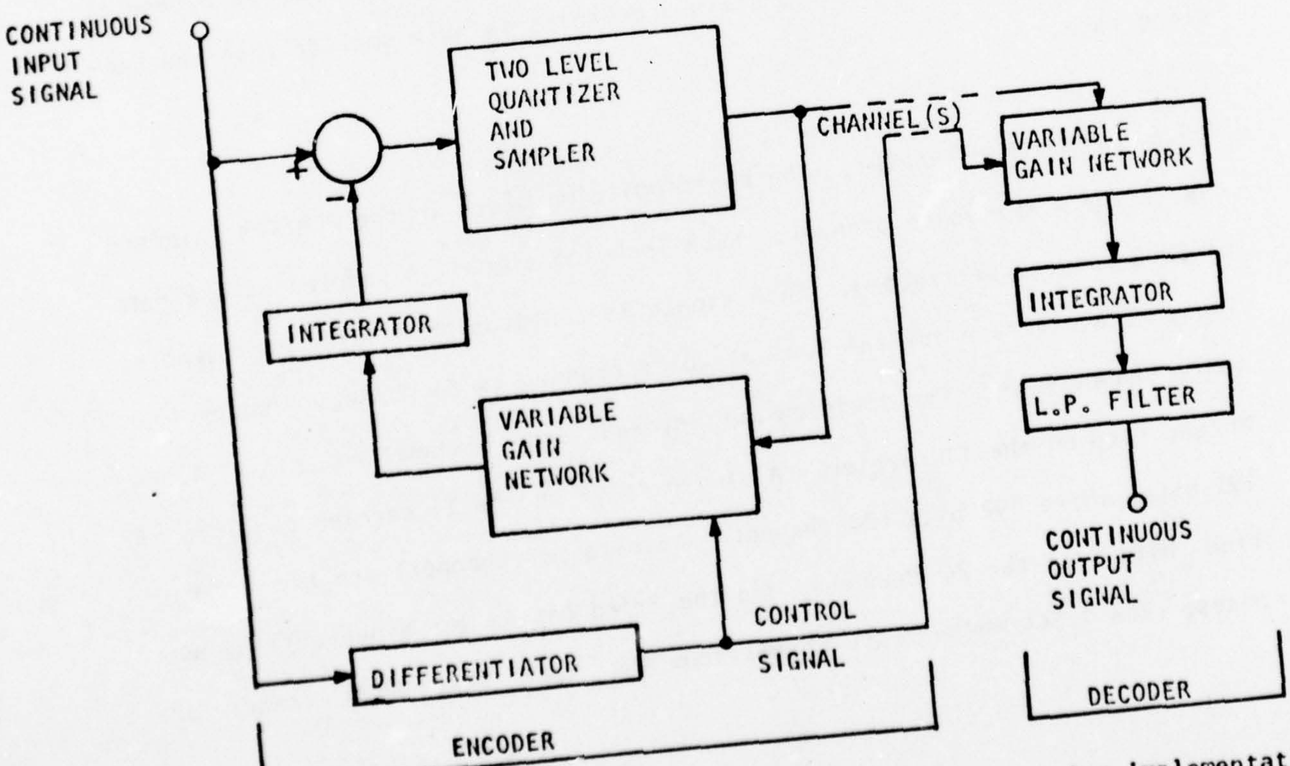


Figure 6-13. Continuous Variable Slope Delta Modulation: analog implementation

same polarity - plus 1's if the signal is increasing and minus 1's if the signal is decreasing. Based on these consecutive samples (usually three in implementation), the switch control selects a new gain which is larger than the previous one, thereby combatting any slope overload. Since the step size is changed at a rate equal to that of the sampling rate, the bit rate requirements remain unchanged.

The continuous adaptive scheme is shown in Figure 6-13. A continuous method of adapting the system to changes in the signal slope is used in this case. The input signal is passed through a differentiator to determine the information regarding its slope. The output from the differentiator is used to control the quantum step size by controlling the gain of the variable gain network. The control signal from the differentiator is also sent to the receiver so that proper decoding can be accomplished. Because the control signal uses up some of the transmission channel capacity, its bandwidth requirements must be a fraction of the input signal bandwidth. As a result, the rate at which the step sizes are varied is much smaller than the sampling rate.

6-9.6. **PCM/TDM HIERARCHY.** The basic building block in the PCM/TDM hierarchy is the 4 KHz voice channel. Each incoming signal is sampled at the rate of 8000 samples per second. Each sample is coded using 8 bits per sample. Thus, the bit rate for the nominal voice channel is 64 Kb/sec. Twenty four voice channels are time division multiplexed and transmitted at a 1.544 Mb/sec rate on the T1 carrier. A single frame on the T1 carrier consists of 193 bits, where 192 bits (24 channels X 8 bits per channel) are the information bits from the 24 channels, and the 193rd bit is for synchronizing purposes. The 1.544 Mb/sec rate arises from the fact that 8000 frames are

transmitted every second--(8000 frames/sec) X (193 bits per frame) = 1.544 Mb/sec. The T1 format has become a standard in the United States and Canada. The CCITT recommendation is different in the number of multiplexed channels: The nominal voice channel is 64 Kb/sec, but 32 channels are multiplexed to give a 2.048 Mb/sec rate. In actual transmission, 30 channels are used for transmitting information from the incoming channels and the remaining two channels are used for synchronizing purposes. The 2.048 Mb/sec rate is commonly used in Europe.

Four T1 pulse streams are multiplexed to form a 6.3 Mb/s pulse stream, which is transmitted over the T2 repeatered line. It should be noted that the T2 line rate is slightly higher than four times the T1 rate; this is because unsynchronized lines can be multiplexed on this carrier system using pulse stuffing, which accounts for the added bit rate.

6-9.7. DATA TRANSMISSION. While PCM systems are designed especially for voice transmission, their digital format makes them particularly good carriers of data. Data modem sets are used to handle slow-speed data, as would occupy no more than one voice channel. The data modem tonal output (if the data set is designed to condition data signals for FDM transmission) is sampled by the PCM channel bank the same way as with voice signal. For higher speed data, special modems work directly into the repeatered line and in some cases are able to coordinate directly with the D channel bank.

There is an advantage of using the PCM line for data transmission. A 50 Kbit/sec data signal displaces 12 voice channels on an FDM system, using 9.6 Kb modems. It takes only one channel on the T1 repeatered line. This efficiency allows for data that is not in synchronization with the line rates: the efficiency is further improved if the incoming data rates pre-

cisely match the bit rates of the incoming lines, because these incoming signals can then be multiplexed on a bit-by-bit basis.

However, when asynchronous data is to be handled, additional treatment is necessary. Three T1 carrier bits are required for each data bit: the first transmitted bit indicates a data transition has occurred; the second bit carries information on the length of the data bit; the final bit relays the direction of transition--that is, plus or minus. Therefore, every time a data bit is received, three successive T1 bits are needed to transmit the information.

When data signals of the type designed for use on analog lines of 4KHz bandwidth encounter a section of digital transmission, one must be careful. The standard 64Kb/s PCM system can handle any such data modem signal. Delta Modulation cannot handle the higher speed modem signals.

6-10. SUMMARY

Our discussion in the previous sections dealt with some of the commonly encountered multiplexing and concentration techniques. Even though there are basic differences in the operation of multiplexers and concentrators, the basic advantage accrued from their use is the same: increase in line efficiency. When used outside the central office (C.O.), both concentrators and multiplexers require fewer physical circuits to the central office than there are subscribers in the area.

Multiplexers provide a trunk to the central office for each incoming voice or data channel by sharing frequency bands (FDM) or time slots (TDM) on a predetermined manner. Both FDM and TDM schemes were presented in this chapter, along with the ramifications of each scheme, such as sub-carrier FDM, asynchronous and synchronous TDM, and statistical time-division multi-

plexing (STDM). Hierarchical multiplexing structures associated with FDM and TDM were also introduced in this chapter. These included: sub-carrier FDM, which combines several low-speed data sources over a single voice channel; voice-carrier FDM, which combines several voice channels to form a wideband transmission facility; low-speed data TDM, which includes 75×2^N bits/s hierarchy.

Concentrators, on the other hand, use a sharing or switching scheme in which some number of input channels share a smaller number of output channels on a demand basis. As a result, it is not possible to have all concentrator subscribers using their equipment simultaneously. The traffic pattern plays an important role in planning and designing concentrators. In dealing with concentrators, concepts regarding multipoint network, speech interpolation, and line concentration were presented.

Finally, PCM/TDM was also presented in this chapter. All important aspects involved in PCM, such as sampling, quantization, and coding were described. Differential pulse code modulation (DPCM), delta modlation, and adaptive delta modulation schemes were also introduced, along with the present-day PCM/TDM hierarchy. In 1978, the Bell System stated that for new voice installations, compared to FDM TDM was more cost effective for all applications with transmission paths less than 50 miles long. With reduced costs of digital equipment, this distance is ever increasing. When data and voice are mixed, the reduced cost of multiplexing and the elimination of quasi analog modems for the data give a greater advantage to digital. The DOD wants to go all digital for security reasons. Thus digital transmission in standard format is becoming the rule.

SECTION VII

NETWORK PLANNING

7-1. GENERAL

The rapid scientific and technological advances coupled with steadily increasing demands for services from users places an important and heavy responsibility on engineers involved in planning networks for transmission. Since the financial investment in telecommunications plant is very large, it is important that the planners be fully aware of certain fundamental considerations needed in the planning process. This section gives a treatment on these considerations. It is emphasized that in attempting to undertake the burdensome task of network planning, one must be fully acquainted with the material of the preceding sections.

7-2. FUNDAMENTAL CONSIDERATIONS

Before embarking on a discussion of the network planning process, it is essential to clearly define and understand the <u>composition</u> and the <u>objectives</u> of a telecommunication network. Here the case of telephone networks is considered; however, the basic principles that are applicable to telephone networks also apply to networks providing other services, such as Telex, data, etc., since most of these other services are carried via the telephone facilities.

7-2.1. NETWORK HIERARCHY. The telephone network is composed of transmission links and switching centers distributed so as to enable every subscriber to make a call to any other subscriber of the system on demand. Much of the traffic originating in a telephone network is of the voice type;

however, services other than telephone services are also provided by the switched telephone network. Economic factors and the historical build-up of the telephone system has led to a hierarchy of networks. This hierarchy comprises:

1. Local networks, which are made up of the subscriber equipment and the local loop connecting the subscriber telephone set to the local switching office.

2. Junction networks, which provide the interconnections between two local offices or between a local office and a trunk office.

3. Trunk networks, which consist of trunk offices and interconnecting transmission trunks.

4. The International network, which, as the name implies, consists of gateway exchanges and interconnecting "trunks" between different countries.

Many advantages are accrued from this hierarchical composition of the telephone system. Some of these advantages are: for operational and analysis purposes, one need only deal with smaller and more manageable units in terms of size and complexity; the modifications in the network can be directly related to changes in local, national or international demands for services; finally, and perhaps most importantly, the hierarchy enables the planner to use a diakoptical approach towards network planning and analysis in contrast to an approach in which the entire network has to be analysed and planned in one piece.

7-2.2. TELECOMMUNICATION NETWORK FUNCTIONS. Next, the fundamental objective of a telecommunication network must be clearly understood and defined. The aim of a switched telecommunication network is to allow any pair of subscribers, irrespective of their relative location, to be able to converse or

send information over a telephone connection with as little difficulty as possible, and at an acceptable cost. This general statement of aim leads to the following basic objectives of a telecommunication system. It must

1. meet the prescribed requirements regarding (a) speed of connection and probability of a busy connection and (b) quality of transmission, which is usually defined in terms of how well the received signal replicates the transmitted signal.

2. uniquely identify subscriber connections.

3. Provide signaling information for call establishment and revenue collection, if needed.

4. Avoid inter-channel interference.

5. be capable of future expansion and modifications.

6. be cost effective.

7-3. NETWORK PLANNING PROCESS

Based on the composition and function of a telecommunication network, it is possible to enunciate the various aspects of the network planning process: the "composition" leads to (1) Local Network planning, (2) Trunk area planning; the functional aspect leads to (1) Strategic planning, and (2) Network Standards. Each of these aspects are discussed in the following paragraphs. But before that, it is important to bear in mind that time scales play an important role in the network planning process. Thus, strategic planning, as mentioned above, should look at 20 to 30 years hence and clearly indicate the trend and direction the network should take, along with the options available during this time period. Similarly, short-term planning may look at the immediate future and make amendments in direct response

to system growth and technological advancement. Time lapse between planning, fabrication, and the completion of the network plays an important role in network planning. The anticipated growth in the type and amount of traffic between the time of the receipt of mission objectives, and its completion, should be taken into account in the original planning process. Otherwise, the network, upon completion, may not meet the standards in terms of the grades of service.

7-3.1. STRATEGIC PLAN. As the name implies, the strategic plan contains a practicable and feasible strategy (or solution) for meeting the present demands, and the anticipated demands in the next decade or more, using the existing technology, and the anticipated changes in technology in the future, in a cost-effective manner. Such planning therefore entails a knowledge of: network modeling and analysis (teletraffic engineering); economic, technological and traffic forecasting; and network standards.

A strategic plan should provide the following information:

1. Traffic Estimate: Estimation of traffic to be carried by the network over the period for which the strategic plan is to be effective is important for sizing and layout of the network. The special services, such as narrowband or wideband data, required should be clearly pointed out in the plan, because of their effect on the planned network performance, etc. Traffic forecasting in adequate detail is required, and should take into consideration such factors as population movements on a local and national level, patterns of industrial activity, and economic growth of the nation. Growth in busy-hour traffic in erlangs on particular routes is primarily determined by local factors; growth in the total traffic originating in the

network depends more on the business activity and other economic indicators, such as the gross national product (GNP). It should be noted that in generating the traffic estimates, both the traffic originating and terminating at a "node" should be considered. Sizing of the trunks relying solely on the originating traffic estimates can lead to poorly designed network in terms of performance.

2. Technological Solutions. Based on the estimated traffic, all technologically feasible solutions that will meet the prescribed requirements must be examined and evaluated subject to cost considerations. New technological solutions likely to be available within the strategy plan period must also be considered.

3. Economic Study: An investment appraisal should be an integral part of a strategic plan. A flow-chart, depicting the principal steps involved in an investment appraisal endeavor, is given in Fig. 7.1. Based on the network objectives called for in the mission statement, a preliminary set of solutions, meeting the stated objectives, is obtained. These are then screened to eliminate the technically unsound and unacceptable alternatives. It is important that all possible solutions are considered before screening. The next step is to make a preliminary cost estimate of the acceptable solutions and to single out, on the basis of the cost estimate, those alternatives that warrant a detailed analysis. It is important to note that economic analysis in network planning is (and should be) rarely done using capital cost, since such an analysis would ignore the time value of money and charges for operation, maintenance and differing plant lives. Discounting methods for evaluating cost which assume that money available now is worth more than an equal sum in the future are used. Two such techniques are: First, the net present value (n.p.v.) or discounted cash flow

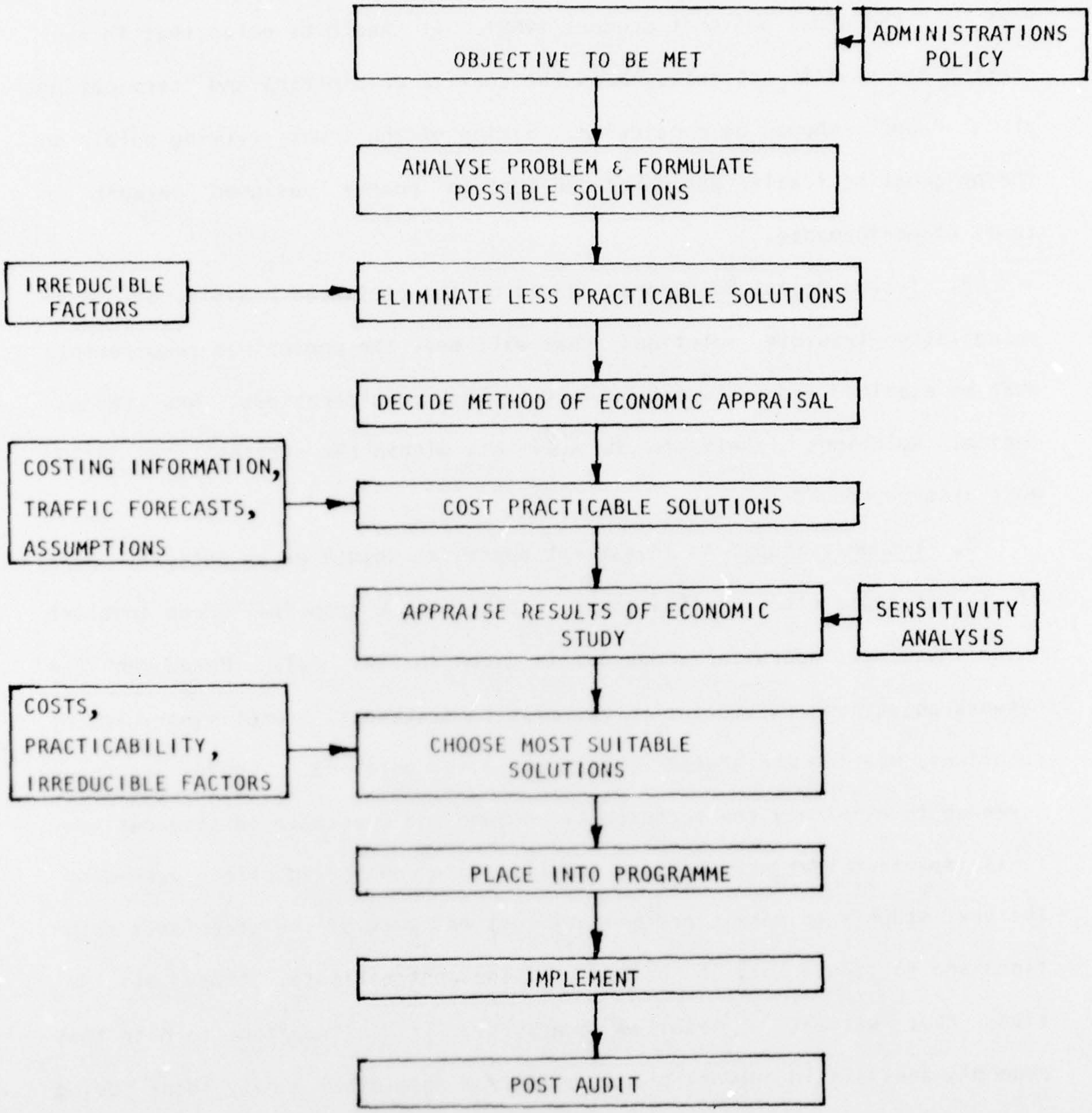


Figure 7.1. Economic Study Flow-Chart.

(d.c.f.) method, along with its variant the internal rate of return (i.r.r.) method; second is the present value of annual charges (p.v.a.c.). The discounting methods basically rely on the method of computing the compound interest on a capital investment to compute the present value of future money.

Note. For a more detailed treatment of investment appraisal, refer to [4].

4. Generalized Objectives: On the basis of the technical and economic analysis of the optimum network meeting the objectives set forth in the mission statement, generalized objectives regarding types of equipment, significant network changes, any major changes (moving from analog to digital, terrestrial to satellite, etc.) along with transition costs, etc. should be clearly brought out. Sensitivity studies and conclusions should also be incorporated within these generalized objectives.

It should be noted that in charting a strategic plan, the following information should be at hand: (1) knowledge of the existing network, (ii) plant details, and (iii) transmission and switching standards.

7-3.2. **NETWORK STANDARDS.** In order to assure a satisfactory operation of the network under planning, various factors that directly affect the network performance must be considered. Due to their importance in network planning and network interfacing, some of theese factors have been standardized. The network planner must be fully acquainted with these standards, which are qualitatively indicated in the following paragraphs.

1. Transmission Plan. Transmission planning involves the switching and trunking aspects of the network. Some of the key factors that must be addressed during transmission planning are:

(a) The allocation of transmission losses and impairments (distortion, added noise, etc.) between local, junction, and trunk portions of the network.

(b) The size of local exchanges, which depends on the number of subscribers, type of message transmission, line sharing, etc.

(c) The switching involved in call completion and such factors as provision for direct or alternate routing, call control rules, maximum number of transmission links to be used in tandem, etc.

2. Routing Plan. The strategy for completing a call between a source and a destination is depicted by a routing plan. The following factors must be considered when devising routing plan:

(a) The transmission plan, which influences the maximum number of tandem links in routing.

(b) Call set-up time, which depends on the type of signaling and switching equipment in use and the mode of transmission (terrestrial vs. satellite transmission).

(c) The economics of direct routing vs. alternate routing.

(d) The acceptable end-to-end (or the node-to-node) grade of service--node-to-node grade of service being defined as the lost load between a source-destination pair divided by the total load between the source-destination pair.

3. Numbering Plan. The numbering plan is used to identify each subscriber

connected to the system. Some basic considerations in devising a numbering plan are:

(a) The average number of digits to be dialed should be minimum and the dialing procedure should be simple.

(b) CCITT recommendations for the international numbering plan.

(c) Allowance for growth in the number of subscribers.

4. Charging Plan. This factor should consider local-call charging, trunk-call (long distance) charging, and international charging.

5. Signaling Standards. Signaling standards cover both supervisory and information signaling. Signaling is an important factor in the network operation; poorly planned signaling can cause excessive call establishment times, which will adversely affect the network performance.

6. Grade of Service. Grade of service is the percentage of calls allowed to be lost due to congestion. The maximum value for the acceptable grade of service in a network is a subjective matter, based on the cost effectiveness of providing service to a subscriber. It is important to distinguish between the various types of grade of service encountered in telecommunication: 1. Node-to-node grade of service, which is the proportion of calls lost due to congestion between two distinct nodes in the network; 2. Link grade of service, which is defined as the calls blocked by a link divided by the calls attempted on the link; 3. Network grade of service, which is the proportion of the total traffic lost in the network due to congestion.

In deciding the grade of service for each link in the network, the function of each link, its cost, and its relative importance in the network should be taken into account.

7-3.3. LOCAL NETWORK PLANNING. Local network planning addresses the details concerned with the size of an exchange area, the location of an exchange, and the exchange(s) (local central office) itself. This aspect of planning may involve the determination of the local-exchange areas and deciding on the optimum location of the local exchanges in an area for which no telephone service exists, or it may involve the modification and augmentation of the existing local network in response to an increase in the number of subscribers requiring services. Local network plans should be reviewed regularly so that alterations in exchange areas, replacement of obsolete equipment, provision of new services, etc. can be carried on in a gradual manner. A local network plan must consider the factors in the following paragraphs.

a. Size of an Exchange. Minimization of cost, while complying with the strategic plan and network standards (presented in the preceding paragraphs), is of paramount concern in network planning. Therefore, the number of local exchanges in a territory and the sizes of their areas must be determined by economic comparisons between different alternatives. A large number of exchanges in a given area means smaller areas for each exchange which, in turn, reduces the length of subscriber loops and their cost. But this increases the number of junctions between exchanges and the number of buildings and sites for the switching equipment. A small number of exchanges (with large exchange areas) reduces the cost of junction plant,

switching equipment, buildings and sites, but results in an increase in the cost of the subscriber distribution networks. This example illustrates the cost trade-off involved, and the need for determining the optimum number of exchanges that yield the minimum cost.

The starting point for determining the boundaries of areas and locations of exchanges must be the anticipated development of the area under consideration and the density map, depicting the population, of the area. Rivers, railways, and roads notwithstanding, the boundary of an exchange should be placed where the density of subscribers is low, to prevent adjacent subscribers being connected to different exchanges. Within the area, the exchange should be located where the telephone density is high, so that the average length of the subscriber loop, and concomitantly the cost of the wire, is reduced. The determination of area boundaries and exchange locations is generally an iterative process.

b. Location of an Exchange. The location of an exchange can be determined after the exchange area has been mapped out. This is usually carried out by a two stage process: First, the theoretical center, which gives the exchange location that would minimize the total length and cost of subscriber lines if there were no existing cables in the area and no topographical features, is found; second, the determination of the practical center, which defines a point in the area at which the exchange should be located to minimize the total cost of the line plant when the existing cable is used. The methods used for determining these two centers do not take into consideration the availability of sites for the exchange. Thus, it often happens that the practical center does not coincide with the available exchange

site. In such cases, the out-of-center cost for each available site is calculated. This is the additional line-plant cost incurred by building the exchange away from the practical center. The out-of-center cost and the site costs are the final arbiters in the selection of an exchange location wherever more than one feasible solution exists. Suitability of sites for cabling and accessability for staff and vehicles are also factors to be considered in an exchange location selection problem.

c. Exchange Planning. Exchanges perform an important function in a telecommunication network. Their functions and status in the network hierarchy must be clearly defined prior to their design. The number of subscribers' lines and their estimated traffic, the junction routes, and the traffic of each route must be determined before the planning of the size of the exchange can be decided. It is important to bear in mind that there is a time lapse between the initial planning and the completion of the network. A broad forecast of ultimate requirements should be at hand when exchange planning is carried out, so that the exchange trunking and equipment layout can accommodate the anticipated growth. The determination of the floor plan, power plant, size of the building, etc. are all part of exchange planning. The choice of a site for an exchange is also critical, and should be done with the view that it can ultimately accommodate a building at last twice as big as that initially required. This is because it is normally much cheaper to extend an exchange on the same site than to develop a new site.

7-3.4. **TRUNK NETWORK PLANNING.** With the existing telephone systems, it is planning for growth, in contrast to basic or long-term planning involving a

major change of a section of the existing network, that is more commonly encountered in practice. The planning activities for the trunk network is divided into two areas: transmission and switching, with signaling being allied to switching planning. These areas are dealt with in connection with basic planning and planning for growth below.

a. Basic Planning. Basic planning for a trunking network is usually long-term, and in compliance with the strategic plan. While the strategic plan gives an overview ("macroscopic") of the required trunking network, the basic plan is a detailed plan, addressing the difficult problems associated with the move from the existing network to that envisaged by the strategic plan. All alternative solutions must be investigated in detail to determine the optimum network configuration and the most economic transition plan. Fundamental to trunk network planning is the selection of the most economic network layout from a set of technically viable solutions. Thus it involves the determination of the number, location and catchment areas of central offices, the traffic routing in the network and its physical realization. The task involves the collection of traffic data relating to the present and future size of the system, together with the costs and parameters of the present and future types of plant. Various network options, meeting the prescribed standards, are analysed in terms of their economics. The most economic solution is then selected. Inherent in most analysis methods are some assumptions. It is prudent to make a sensitivity study on the preferred network configuration due to variations in the basic data and assumptions used.

The determining factors for the determination of the number, location and catchment areas of trunk exchanges are similar to those for local exchanges. Once the location and catchment areas of trunk exchanges have been determined, the next task is to find the dimension and cost of the total network. The essential part of dimensioning is the determination of the most economic routing plan for the given traffic. This can be done either by simulation studies or analysis methods (single- or double-moment methods). A traffic route matrix of circuits between all switching points is obtained, from which the trunking facility and the switching facility costs can be determined. This planning procedure is repeated for different network configurations. The optimum network configuration is the one that has the least discounted cost.

b. Growth Planning. In the changing environment of a telecommunication network, where both traffic needs and plant availability change with time, growth planning is more common than basic planning, although planning for growth may require some of the basic steps required in basic planning. Forecasts of traffic parameters, usually extrapolated from the current measured traffic, play the key role in growth planning.

Based on the new traffic information for the network, the forecast requirements for each parameter in the network can be determined. These requirements are then compared with the existing network call carrying capacity and other parameters to ascertain when the grade of service and other network standards will fall below the acceptable values. The earliest date indicates the service data required for the next "network" extension. A similar forecasting scheme can be used to determine when the present central

office accomodations will be exhausted. Before carrying out any network extension in response to an increased traffic, it is important that one considers the possibility of re-routing (updating of the existing route plan) the traffic on the existing network to meet the prescribed network standards.

REFERENCES

1. Transmission Systems for Communications, Western Electric Comapny, Inc., December 1971.

2. Carlson, A.B., Communication Systems: An Introduction to Signals and Noise in Electrical Communications, McGraw-Hill Book Company, 1968.

3. Freeman, R. L., Telecommunication Transmission Handbook, John Wiley and Sons, 1975.

4. Flood, J. E., ed., Telecommunication Networks, Peter Peregrinus Ltd. (on behalf of IEE), 1975.

5. Bylanski, P. and Ingram, D. G. W., Digital Transmission Systems, Peter Peregrinus Ltd. (on behalf of IEE), 1976.

6. Bennett, R. and Davey J., Data Transmission, McGraw-Hill Book Company, 1965.